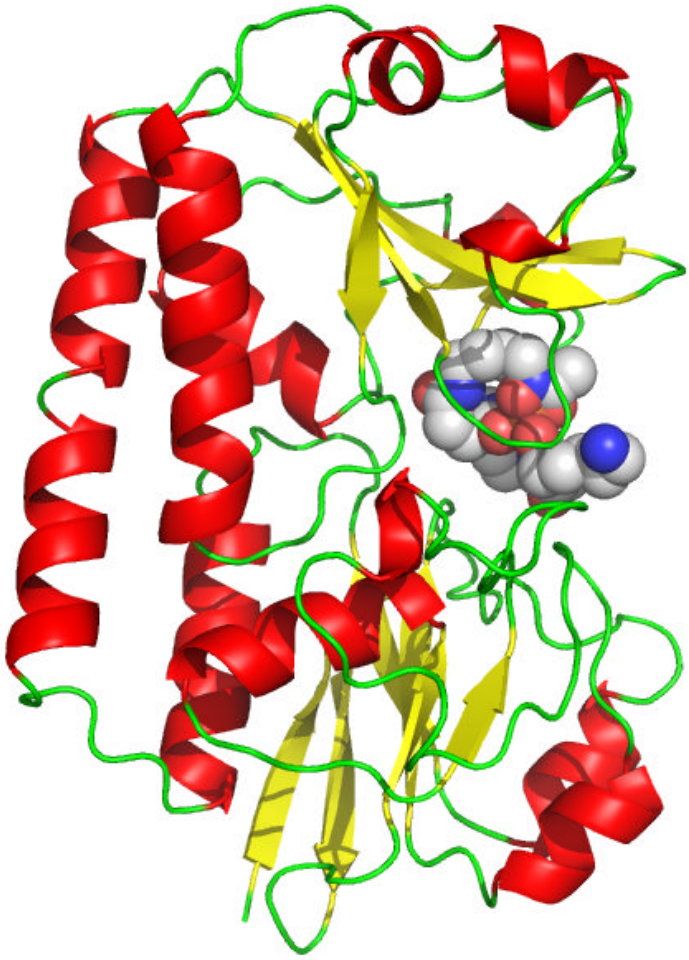
Investigations of *Streptomyces coelicolor* A3(2) siderophore binding proteins



Doctor of Philosophy in Interdisciplinary Science

Name:	Prakash Patel
Supervisors:	Professor Gregory L. Challis
	Professor Robert. J. Deeth
Institution:	University of Warwick
Department:	MOAC
Date:	17th July 2009
University No.:	00008758

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Interdisciplinary Science

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first and foremost thank my supervisor, Professor Gregory Challis, for giving me the opportunity to work in his group on this exciting project and his invaluable guidance, and also his relentless patience regarding my lack of experience in chemical biology. I am eternally grateful to Dr. Xujia Hu of State Key Laboratory of Phytochemistry and Plant Resources in West China, for her assistance in the purification of gallium-coelichelin, and Professor Jim Naismith and his group of the University of St. Andrews for solving the X-ray crystal structures of apo-DesE and DesE in complex with ferrioxamine B. I would like to thank the Challis group, past and present, who have been a pleasure to work with. Special thanks to Dr. Nadia Kadi, Dr. Christophe Corre and Dr. Lijiang Song for their assistance in this project. I would also like to thank Professor Alison Rodger and her group for their warm hospitality for the few months I stayed there to complete my experiments.

I would also like to thank my co-supervisor, Professor Robert Deeth, for a very pleasant year as a computational chemist, and also the members of his group, both past and present: in particular Kris Randell and Dr. James Burnside, both of whom I will look up to always!

There are three amazing individuals I would like to thank for one of the greatest achievements my life: my three victorious University Challenge team-mates Rory Gill, Harold Wyber and our scintillating captain Daisy Christodoulou. The whole experience was exhilarating to say the least, resulting in the first University Challenge series win for Warwick. These three people have made me realize how little I knew about the world, and possibly science as well. I would also like to acknowledge Jeremy Paxman and the Rt. Hon. Ann Widdecombe MP for their roles in the proceedings.

Finally, I would like to thank my friends and colleagues at in the Chemical Biology Cluster, and all the students and staff at the MOAC Doctoral Training Centre, especially the 2004 intake. There are a few others I would like to thank, in particular Bugadoo bear, our University Challenge mascot, for his mere existence on this planet. I would like to thank my family for their constant support and

encouragement throughout the 3½ years, in particular my brother Rakesh, my sister Vina and my brother-in-law Nilesh. This thesis is dedicated to my parents, Tara and Dhiru. The Engineering and Physical Sciences Research Council is thanked for their financial support.

DECLARATION

Experimental work contained in this thesis is original research carried out by the author, unless otherwise stated, in the Department of Chemistry and the MOAC Doctoral Training Centre at the University of Warwick, between October 2005 and December 2008. No material contained herein has been submitted for any other degree, or at any other institution.

References from other authors are referenced in the usual manner throughout the text.

_____ Date: _____

Prakash Patel

ABSTRACT

Siderophores are small, high-affinity ferric iron chelators released by many microorganisms and some plants to solubilize iron. They are of great interest due to their clinical use to treat iron overload in humans, and also in relation to the development of novel antibiotics that target the biosynthetic and uptake pathways for iron in pathogens. Pathogens such as *Bacillus anthracis* excrete more than one type of siderophore. This is linked to increased pathogenicity. The Gram-positive soil bacterium *Streptomyces coelicolor* A3(2) excretes three siderophores: desferrioxamine B, desferrioxamine E and coelichelin. These displace iron from insoluble ferric hydroxides, and the resulting ferric complexes are transported into the cell via siderophore-binding proteins (lipoprotein receptors) associated with ATP-binding cassette (ABC) transporters. Previous studies showed that some of the genes in the biosynthetic clusters of the desferrioxamines (*des*) and coelichelin (*cch*) were required for efficient uptake of ferrioxamine E and ferri-coelichelin respectively and a third ABC transporter gene cluster (*cdt*), not associated with siderophore biosynthesis genes, was implicated in the import of ferrioxamine B.

In this study, the lipoprotein receptors encoded within the *des*, *cch* and *cdt* clusters - DesE, CchF and CdtB – were recombinantly overproduced in *E. coli* and purified by immobilized metal affinity chromatography. Also, ferri-coelichelin was purified from cultures of *S. coelicolor*. The binding of the ferric complexes of the three cognate siderophores, as well as the xenosiderophores ferrichrome and ferri-albomycin, to the lipoprotein receptors was monitored by intrinsic fluorescence quenching. Dissociation constants of receptor-siderophore complexes were found to be in the nanomolar range, and a revised model of cognate siderophore transport in *S. coelicolor* was proposed. In collaboration with researchers at St. Andrews University, an X-ray crystal structure was solved for apo-DesE and DesE bound to ferrioxamine B, which demonstrated the similarity of DesE to other bacterial siderophore-binding proteins and the negligible conformational change on substrate binding. Ferrioxamine B also exhibited an unusual configuration not observed before in X-ray crystals of this ferri-siderophore. Also, a forcefield was constructed to model the structure and distortions ferric-*tris*-hydroxamate complexes, which

could be used in the future to investigate the molecular basis of the tight and specific binding of ferri-siderophores to siderophore-binding proteins.

ABBREVIATIONS

AAS	Atomic absorption spectroscopy
ABC	ATP-binding cassette
ADF	Amsterdam Density Functional
ALB	Ferri-albomycin
AOM	Angular Overlap Model
ATP	Adenosine triphosphate
ATPase	ATP hydrolase
bps	Base pairs (DNA)
BSA	Bovine Serum Albumin
CCLIN	Ferri-coelichelin
CCSD	Cambridge Crystallographic Structural Database
CD	Circular Dichroism
CDS	Chemical Database Service
CM	Cytoplasmic Membrane
COSMO	Conductor-like screening model
COSY	Correlation Spectroscopy
Da	Daltons
DFT	Density Function Theory
DmdR	Divalent metal-dependent regulatory protein
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DtxR	Diphtheria toxin repressor
EDTA	Ethylenediaminetetraacetic acid
Ent	Enterobactin
ESI	Electrospray
fac	Facial isomer
FAD	Flavin Adenine Dinucleotide
FBP	Ferric binding protein
FCR	Ferrichrome
$Fe(hydr)_3$	Ferric tris-(cis-acetohydroxamic acid)
$Fe(Nmhydr)_3$	Ferric tris-(cis-N-methyl-acetohydroxamic acid)
FF	Forcefield

FOB	Ferrioxamine B
FOE	Ferrioxamine E
GC	Guanine-Cytosine
GF	Gel Filtration
HF	Hartree-Fock
His_6	Hexahistidine tag
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear spin quantum coherence
ICP-OES	Inductively coupled plasma optical emission spectroscopy
IMAC	Immobilized metal-affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
kpbs	Kilo base pairs (DNA)
LB	Luria-Bertani Broth
LDA	Local Density Approximation
LE	Low electroendosmosis
LFSE	Ligand-field stabilization energy
MALDI	Matrix-assisted laser desorption ionization
MBP	Maltose binding protein
MCMC	Markov Chain Monte Carlo
MECAM	N”N”-tris(2,3- dihydroxybenzoyl)-l,3,5-tris(aminoethyl)benzene
MeCN	Acetonitrile
mer	Meridional isomer
MM	Molecular Model
MOE	Molecular Operating Environment
MP2	Møller-Presset peturbation
MS	Mass spectrometry
MSD	Mean squared deviation
MW	Molecular weight
NBD	Nucleotide binding domain
NCBI	National Center for Biotechnology Information
NIS	NRPS-independent siderophore synthetase
NMR	Nuclear Magnetic Resonance

NPRS	Non-ribosomal peptide synthetase
NTA	Nitriloacetic acid
OM	Outer Membrane
OMR	Outer Membrane Receptor
OVA	Ovalbumin
PAGE	Polyacrylamide Gel Electrophoresis
PBP	Periplasmic binding protein
PBS	Phosphate-buffered saline solution
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PMSF	Phenylmethylsulphonylfluoride
PW91	Perdew-Wang 91 (gradient corrected functional)
QM	Quantum mechanics
RBS	Ribosome binding site
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
RP-HPLC	Reverse-phase HPLC
rpm	Revolutions per minute
SBP	Siderophore binding protein
SCF	Self-consistent field
SDS	Sodium Dodecyl Sulfate
SFM	Soya Flour Mannitol medium
SIP	Siderophore interacting protein
SVL	Scientific Vector Language
Taq	*Thermus aquaticus* (polymerase)
TAT	Twin-arginine translocase
TBE	Tris-boric acid EDTA buffer
TCI	Triple Resonance Inverse
TEV	Tobacco Etch Virus (protease)
TFA	Trifluoroacetic acid
TMD	Transmembrane domain
TOF	Time-of-flight
Tris	Tris(hydroxymethyl)aminomethane

UV	Ultraviolet
XAFS	Extended X-Ray Absorption Fine Structure Spectroscopy

Chapter 1: Introduction

1.1 : Iron in living systems

Iron is essential for almost all life, being the active constituent of many redox enzymes such as haem-containing proteins, catalases, peroxidases and iron-sulfur proteins[1], essential for processes such as respiration and DNA biosynthesis. Some lactic acid bacteria are an exception: manganese and cobalt are used instead of iron[2].

Despite the fact that iron is one of the most abundant elements in the Earth's crust, many microorganisms and plants have difficulty obtaining enough iron in non-acidic, oxygenated environments as it is usually found in the trivalent form Fe^{3+}, which forms hydrated hydroxides (such as rust) that are insoluble and notoriously stable[1]. By contrast, the Fe^{2+} ion is soluble but is invariably oxidised via the Fenton reaction by hydrogen peroxide in aerobic conditions[3]:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{\cdot} + OH^-$$

In neutral or alkaline environments, the concentration of free Fe^{3+} ions in solution can be as low as 10^{-18} mol L^{-1} at neutral pH[1]; this is far below the physiological concentration needed for microorganisms to survive.

The scarcity of iron in animals is also a fundamental issue for pathogenic bacteria. In mammalian hosts, iron is tightly bound to proteins such as haemoglobin, the iron transport protein transferrin, defence protein lactoferrin and the iron storage protein ferritin. The strict homeostasis of iron leads to a free concentration[4] of about 10^{-24} mol L^{-1}, hence there are great evolutionary pressures put on pathogenic bacteria to obtain this metal.

1.2 : Iron acquisition by microorganisms

To obtain enough iron to survive and proliferate, both prokaryotes and eukaryotes have developed mechanisms to obtain iron from the extracellular environment.

There are two main ways this is achieved: the direct binding and extraction of iron from extracellular proteins, and the indirect extraction of iron from other extracellular sources.

1.2.1 : Direct binding and extraction from proteins

Pathogenic Gram-negative bacteria such as *Neisseria* and *Haemophilus* have outer membrane transporters that bind to the host's transferrin and lactoferrin[3,5]. As these proteins are too large to pass through the bacterial outer membrane, the iron carried by these proteins has to be extracted, probably by the separation of the domains containing the iron-binding sites. The free Fe(III) ion in the periplasm is captured by ferric ion binding proteins (FBP)[6], and the iron contained within is shuttled to the cytoplasm via adenine triphosphate (ATP)-binding cassette (ABC) protein transporters[3].

Some bacteria also bind to haem-containing proteins such as haemoglobin, myoglobin and the haem-sequestration proteins haemopexin and albumin. Instead of ferric ion being transported, the haem group is extracted from the haem protein and transported into the cell[7]. The Gram-positive bacterium *Staphylococcus aureus* contains a haem uptake system consisting of three cell-surface lipoprotein receptors that bind to different haem proteins, which in turn are substrates for an ABC transporter system, which extracts and transports only the haem groups into the cytoplasm[8]. Free haem is reactive and cytotoxic, hence it is degraded to release the ferrous or ferric iron[7].

1.2.2 : Indirect extraction of iron from extracellular sources

Microbes can increase the concentration of extracellular soluble iron by releasing small molecules that scavenge ferric ion from ferric hydroxides and iron transport proteins. The resulting ferric complexes are actively transported into cells via specific transporters, or are reduced extracellularly to release the iron from the ferric complex. These small molecule iron scavengers are known as **siderophores** (Greek: "*iron carrier*") and is by far the most widespread means used to obtain iron amongst the kingdoms of life[9].

By analogy, some organisms can scavenge haem from haem proteins via small proteins called haemophores[10]; but only the haem is transported into the cells as the haemophore-haem complex is too large for absorption. Analogous processes are used to transport free haem groups as with ferric-siderophore complexes into cells. The weakness of haemophores is that they are of course restricted to haem-containing iron sources, whereas siderophores can sequester iron from almost all iron sources[9].

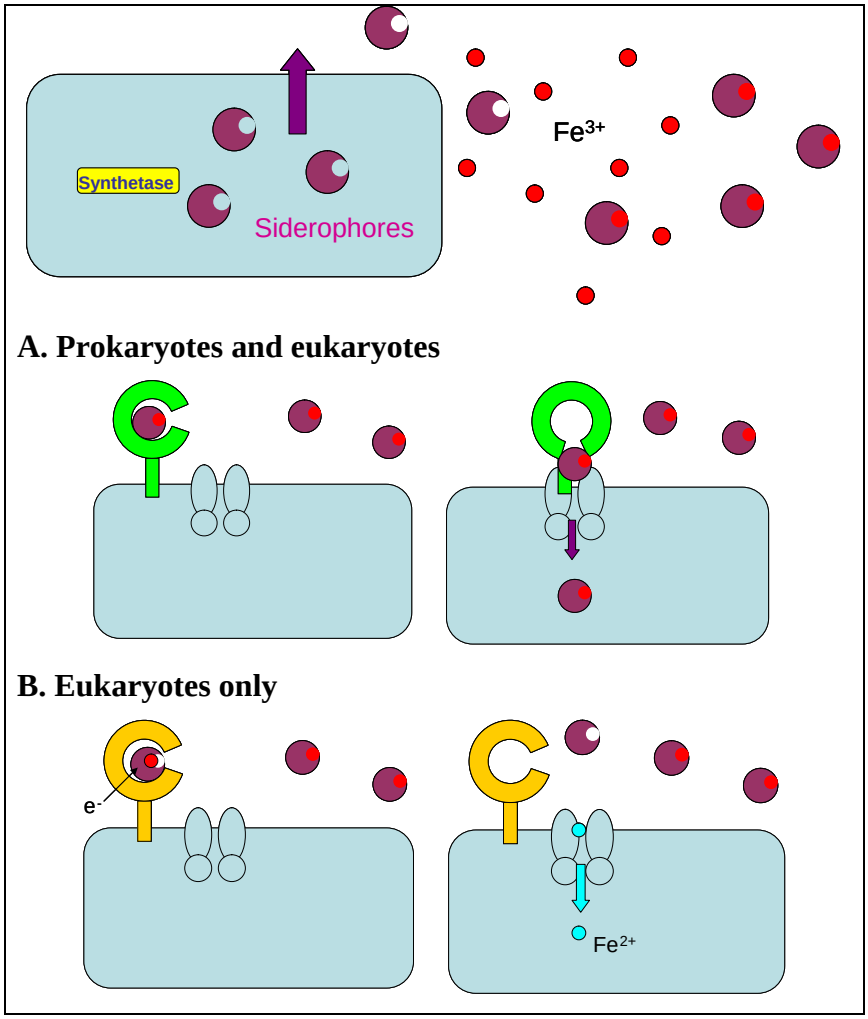


Figure 1-1: Siderophore-mediated iron acquisition.

Under iron deficiency, iron dependent repressors derepress to allow transcription of genes involved in siderophore biosynthesis and uptake. Synthetases manufacture siderophores (purple crescents) that are exported to the extracellular medium where they bind to ferric iron (red). A. Once the cell encounters the Fe-siderophore complexes, in most organisms they are recognised by high-affinity membrane-associated receptors that transport the complete Fe-siderophore complex into the cell to release the iron contained. B. Eukaryotes, such as fungi, possess extracellular reductases that can release iron from the siderophore by reduction of the iron centre. The Fe(II) ion can be transported into the cell via specific Fe(II) importers.

1.3 : Siderophores

1.3.1 : Introduction

Siderophores are low molecular weight, high affinity iron chelators which are produced by many microorganisms and some higher plants to scavenge iron from precipitates and iron-containing proteins. They are produced in response to intracellular iron limitation and excreted. They bind mainly to ferric iron, and the resulting ferric-siderophore complexes are shuttled back into cells via active transport mechanisms.

Siderophores are amongst the highest affinity Fe(III) chelators known; ferric enterobactin has a dissociation constant[4] of 10^{-52} mol L^{-1}. Their prodigious affinity for metals is not restricted to ferric iron: they have strong affinities for other cations such as Al(III), Ga(III) and Pu(IV)[11]. For this reason they have attracted the interest of medical scientists for iron sequestration in man, but many are toxic or are unstable in the body. The siderophore desferrioxamine B (as the mesylate salt Desferal®), produced by many members of the bacterial genus *Streptomyces*[12], has gained widespread use as treatment for iron overload in humans. Desferal has been a success in treating iron poisoning resulting from overdoses and thalasaemias, despite being mildly toxic and the requirement for several hours of intravenous administration[13].

1.3.2 : Nomenclature

There is much confusion in the naming of siderophores and their ferric (or other metal) complexes. For enterobactin, the ferric complex is referred to as ferric-enterobactin. The ferric complex of desferrioxamine B is known as ferrioxamine B (the des implies the absence of iron). This analogy is extended to the other desferrioxamines. The deferrated form of ferrichrome is known as desferri-ferrichrome, whilst its gallium(III) complex is commonly referred to as gallichrome.

1.3.3 : The structures of siderophores

The molecular mass of siderophores varies from 600 to 1600 Da. They are biosynthesised by both prokaryotes and eukaryotes either using non ribosomal peptide synthetases (NRPS)[14] or NRPS-independent siderophore synthetases (NIS)[15]. Extracellular ferric iron binds to the ligating atoms of the siderophore to form stable (in most cases, hexadentate) complexes.

Fe(III) prefers strong Lewis bases as ligands. These are usually neutral or anionic oxygen electron pair donors[9]. The ligating groups within siderophores are usually hydroxamate, catecholate (phenolate), or α-hydroxy carboxylates such as citrate (Table 1-1)[16] . Citric acid itself can act as a siderophore; bacteria such as *Escherichia coli* have specific uptake systems for ferric dicitrate[17].

Due to the extreme rarity of soluble iron in many environments, microbes need to take advantage of any iron source encountered. Many bacteria produce no more than two types of siderophore, but can utilize many different ferric-siderophores as sources of iron. For example, *E. coli* produce only the siderophores enterobactin and aerobactin, but can utilize the xenosiderophores ferrichrome and ferric coprogen as sources of iron as well as ferric-aerobactin, ferric-enterobactin and ferric dicitrate[11]. The remarkable diversity in the structures of siderophores may be due to some organisms developing siderophores whose ferric complexes cannot be taken up by other bacteria[18], or cannot be recognised by the immune system in the case of pathogenic bacteria[19].

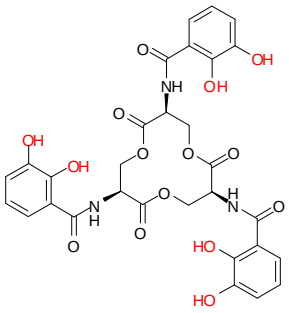
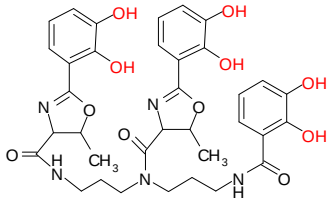
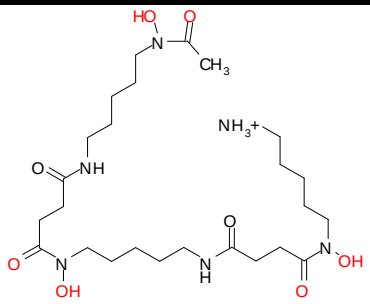
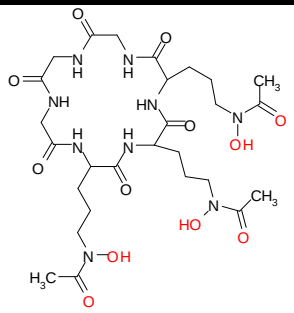
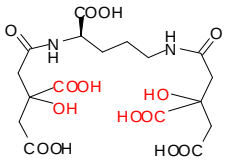
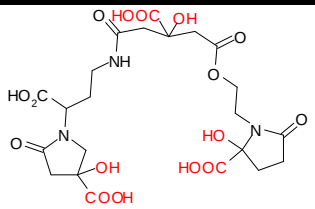
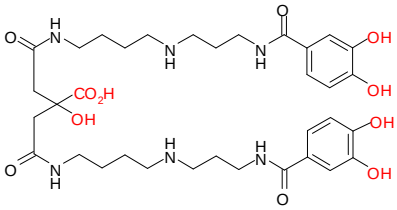
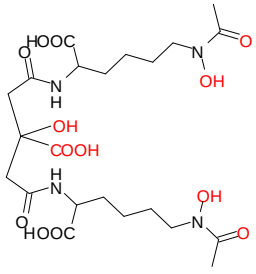
1.4 : Siderophore-mediated iron acquisition

1.4.1 : The regulation of the biosynthesis of siderophores

Microbes need to tightly regulate production of biosynthetic enzymes and transport systems to allow efficient siderophore biosynthesis and secretion, as well as ferric-siderophore uptake and iron release[9].

Table 1-1: The structures of representative siderophores, and examples of organisms that produce them in response to iron deficiency.

The ferric iron binding groups in each molecule are highlighted in red.

Catecholate siderophores	
	
Enterobactin (*Escherichia coli*)	Vibriobactin (*Vibrio cholerae*)
Hydroxamate siderophores	
	
Desferrioxamine B, Desferal® (*Streptomyces pilosus*, *Streptomyces coelicolor*)	Desferri-ferrichrome (*Ustilago sphaerogena*)
α-hydroxy carboxylate siderophores	
	
Staphyloferrin (*Staphylococcus hyicus*)	Achromobactin (*Dickeya dadantii*)
Mixed ligand siderophores	
	
Petrobactin (*Bacillus anthracis*)	Aerobactin (*Escherichia coli*)

The first step in siderophore-mediated iron acquisition is the manufacture of siderophores. Bacterial siderophore utilization and iron homeostasis is usually mediated at the transcriptional level by the ferric uptake repressor Fur in Gram-negative bacteria and low guanine-cytosine (GC) content Gram-positive bacteria, whilst in high-GC content Gram-positive bacteria this role is taken by proteins related to the diphtheria toxin repressor DtxR[9]. These repressors are regulated by the concentration of Fe^{2+} in the cytosol: when Fe^{2+} is bound to one of these proteins, the protein has a strong affinity for a palindromic sequence of DNA bases and transcription of the downstream DNA is repressed. When Fe^{2+} dissociates from its complex with the repressor, in turn, the repressor disassociates from the DNA, allowing transcription of the genes[20]. Searching genomes for the palindromic repressor binding sequence can allow identification of genes downstream which may have a role in the maintenance of iron homeostasis.

1.4.2 : The secretion of siderophores

Siderophore secretion systems have been identified in only a few microbes so far[3]. A recently characterized pathway in *Mycobacterium tuberculosis* involves one protein, IrtA, as a complete ATP-binding cassette (ABC) transporter involved in the export of carboxymycobactin[21].

1.4.3 : Uptake of ferric-siderophore complexes in bacteria

Once the ferric siderophore complex has formed in the extracellular milieu, the ferric ion remains solubilized until a microbe encounters the complex. The ferri-siderophore complexes are too large to be transported by diffusion or via porins. Eukaryotes such as fungi are able to reduce the ferri-siderophore complex to release Fe(II), which can be readily absorbed, because some possess extracellular ferric reductases. In the case of prokaryotes, the ferri-siderophore complex is actively transported into the cytoplasm, where the complex can be reduced or degraded to release the iron (Figure 1-2).

In Gram-negative bacteria, which have an outer membrane and an inner (periplasmic) membrane, the Fe-siderophore complex is taken up via outer membrane receptors (OMRs) and ABC transporter systems, respectively. In Gram-positive bacteria, with only an outer membrane, ABC transporter systems analogous to those in the periplasmic membrane of Gram-negative bacteria import the complex[3].

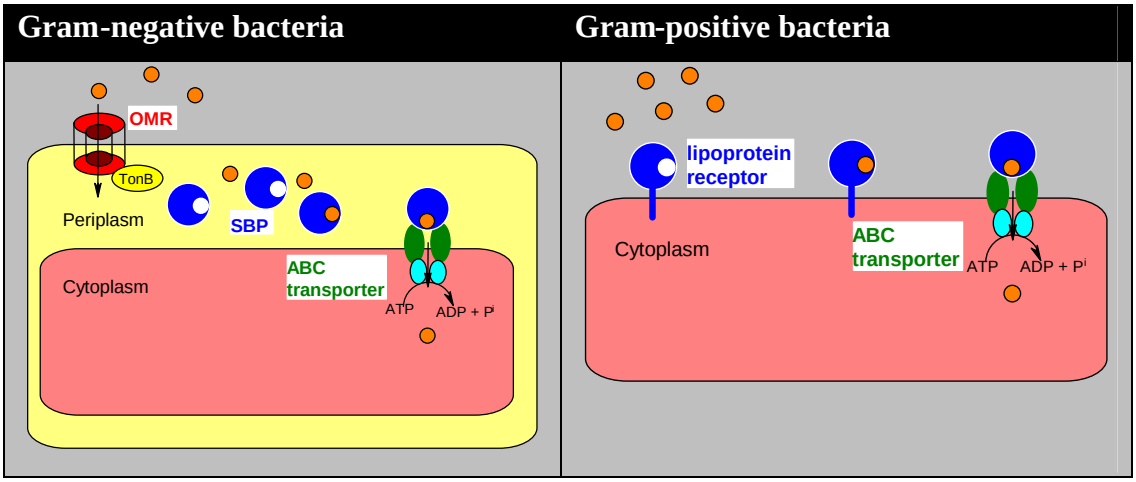


Figure 1-2: General mechanisms of ferri-siderophore uptake in Gram-negative and Gram positive bacteria.

(Left) In Gram-negative bacteria, the ferri-siderophores (orange circles) are taken up by high affinity β-barrel membrane proteins known as outer membrane receptors (OMRs). Active transport is powered by changes in the conformation of the TonB protein. Once in the periplasm, the ferri-siderophore binds to periplasmic binding proteins or siderophore-binding proteins (SBPs), part of ABC transporter systems (permease domains in green, ATPase domains in blue) which transport the ferri-siderophore to the cytoplasm. (Right) Gram-postitive bacteria have only one membrane through which the complex has to pass to enter the cytoplasm. This is mediated by "lipoprotein receptors", tethered analogues to the SBPs, which feed the substrate to analogous ABC transporter systems.

1.4.4 : Outer membrane ferric-siderophore transport in Gram-negative bacteria

The import of ferric siderophores across the outer membrane of Gram-negative bacteria is mediated by OMRs, large β-barrel proteins which span the membrane. The loops between the β-strands of the barrels that are exposed to the extracellular environment are involved in Fe-siderophore recognition (Figure 1-3). X-ray

crystallographic structures have been solved for the *E. coli* OMRs FepA[22] (ferric enterobactin receptor) and FhuA[23] (ferric hydroxamate uptake receptor). They have a very high affinity for their substrates, Fe-enterobactin having a dissociation constant of 20 nM from FepA[9].

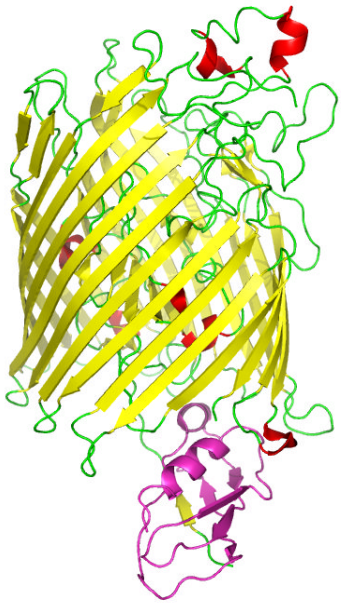


Figure 1-3: Outer membrane transporter *E. coli* FhuA in complex with TonB.

The α-helices, β-sheets and loops of FhuA are coloured red, yellow and green respectively, and TonB is coloured purple. The extracellular loops at the top are responsible for ferri-siderophore binding. PDB code: 2GRX.

The active transport of the Fe-siderophore complex via the OMRs is driven by the changes in conformation of the TonB-ExbB-ExbD complex[24], where TonB interacts with the OMRs whilst ExbB and ExbD reside on the periplasmic membrane. The conformational change is driven by the proton motive force difference between the periplasm and the cytoplasm[24].

1.4.5 : Periplasmic membrane ferric siderophore transport in Gram-negative bacteria

Once the Fe-siderophore is in the periplasmic space, it is transported to the cytoplasm via ATP-binding cassette (ABC) transporters. These consist of periplasmic binding proteins (PBP) or siderophore-binding proteins (SBP), a pair of

permeases (which span the width of the membrane and form a heterodimer) and a pair of ATPases on the cytoplasmic side – the site of hydrolysis of ATP that drive the translocation of the ferri-siderophore[25].

One of the best characterized ABC transporter systems is the *E. coli* BtuCDF system[26], which is responsible for the transport of vitamin B_{12} into the cytoplasm. Vitamin B_{12} is an essential vitamin (a cobalt-corrin complex) that shares many similarities to ferric siderophore complexes in terms of size and shape. The fold of the PBP BtuF is very similar to those of SBPs such as the *E. coli* ferric hydroxamate transporter FhuD (Figure 1-4). Thus, this system is a good model to elucidate the mechanisms of ferric siderophore uptake.

BtuF is a bilobal, "kidney-shaped" protein, consisting of two antiparallel β-sheet regions connected by a rigid backbone of several α-helices. Opposite this backbone is a cavity whose shape is defined by the residues on the extracellular loops of both antiparallel β-sheet regions. An X-ray crystal structure of BtuF in complex with its substrate vitamin B_{12} showed little conformational change compared to the apo protein form[27]. A similar shape and rigidity of the complex is seen in all SBP proteins[26,28] crystallized to date, in particular the *E. coli* proteins FepB (ferric-enterobactin receptor) and FhuD. This is not the case for all periplasmic binding proteins of ABC transporters - the most notable example being maltose binding protein (MBP)[29], whose fold is dramatically altered on substrate binding.

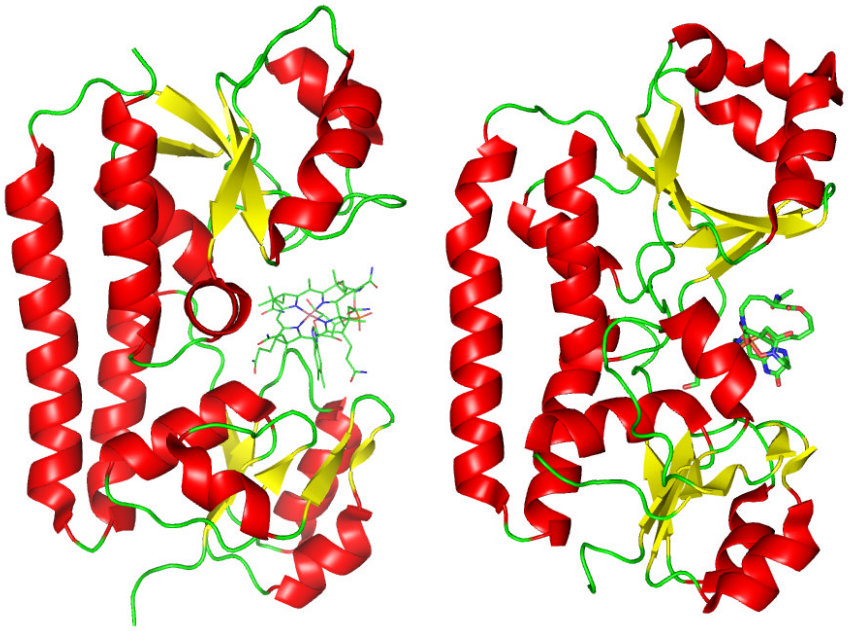


Figure 1-4: X-ray crystallographic structures of the *E. coli* proteins BtuF and FhuD with their substrates bound.

Left: BtuF with vitamin B_{12} bound (PDB code: 1N4A). Right: FhuD with ferric coprogen bound (PDB code: 1ESZ)

The *E. coli* SBP FhuD[30] showed markedly lower affinity for its substrates than the corresponding OMRs, having 10- to 100-fold lower affinity for ferrichrome than FhuA[31]. This is unsurprising as the accumulation of the substrate into the periplasm has already been performed by the OMRs, hence there is little evolutionary pressure to increase the SBP's affinity for siderophores. Ferri-enterobactin has a dissociation constant of 30 nM from the *E. coli* SBP FepB[32]; this is comparable to the OMR FepA.

1.4.6 : PBP interaction with the permease-ATPase domains

A crystal structure has been solved for BtuF in complex with the $BtuC_2D_2$ permease-ATPase dimer (Figure 1-5), but without the substrate. This structure may represent the complex after vitamin B_{12} has been transported[26]. It showed the PBP interacting with the permease dimer $BtuC_2$ via the substrate binding pocket, which also may provide evidence that the vitamin itself may be involved in recognition by the permease domains. Closer analysis of the structure showed the PBP was pulled apart slightly on binding to the permease (via aspartate residues conserved in many

siderophore PBPs) compared to the apo and ligand bound forms of BtuF. This is likely to be the mechanism of siderophore release from BtuF into the $BtuC_2$ pore. Comparisons to a crystal structure of the $BtuC_2D_2$ permease-ATPase dimer[33] showed differences in the arrangement of transmembrane domains of the permease dimer, demonstrating the conformational changes involved in vitamin B_{12} transport.

1.4.7 : Ferri-siderophore uptake in Gram-positive bacteria

Gram-positive bacteria have only one outer membrane (surrounded by a thick peptodiglycan layer) which Fe-siderophores have to pass through to get to the cytoplasm. This is mediated by ABC transporters which are analogous to the ABC transporters of Gram-negative bacteria, except that the SBPs are tethered to the outer membrane by an N-terminal lipid, and are consequently known as "lipoprotein receptors". The X-ray crystal structure of FeuA (Figure 1-6), a ferric-bacillibactin SBP of *Bacillus subtilis*[34,35], confirmed the similarity of these proteins to Gram-negative SBPs. These tethered SBPs are also seen in Gram-negative organisms: for example the enterobactin SBP CeuA[36,37] of *Campylobacter jejuni*. Despite the tether, they are assumed to interact with the permease-ATPase parts of their ABC transporters in a similar way to untethered periplasmic SBPs.

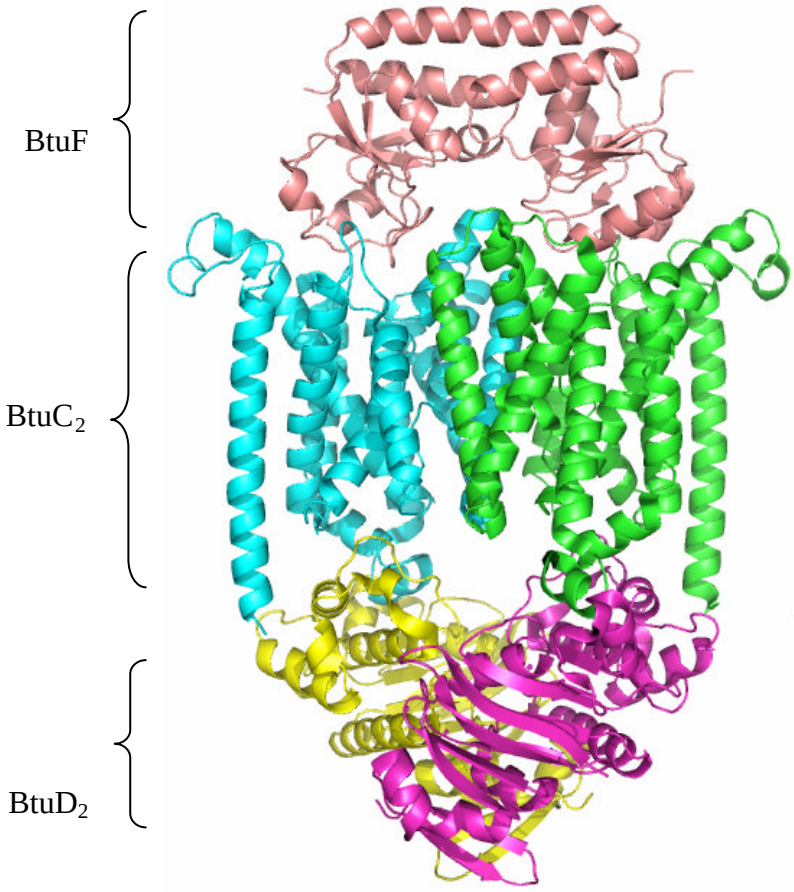


Figure 1-5: The X-ray crystallographic structure of BtuF in complex with BtuC-BtuD dimer.

The periplasmic binding protein BtuF (red) interacts on the substrate binding side with the permease dimer BtuC (blue and green) embedded in the cytoplasmic membrane. The dimer of BtuD ATPases, found on the cytoplasmic side, is coloured yellow and purple. The absence of the substrate vitamin B_{12} from this structure led to the conclusion that this was a post-translocation intermediate (PDB code: 2QI9).

These receptors have higher specificities towards their Fe-siderophore substrates than SBPs of Gram-negative bacteria. Dissociation constants for Gram-positive SBPs are about 2 orders or magnitude smaller than for Gram-negative SBPs, close to those of the OMRs in Gram-negative bacteria. For example, the *S. aureus* lipoprotein receptor FhuD2 has dissociation constants of 20 nM and 50 nM from ferrichrome and ferrioxamine B respectively[38] (compared with *E. coli* FhuD[30], whose dissociation constants are 600 nM and 35 μM respectively).

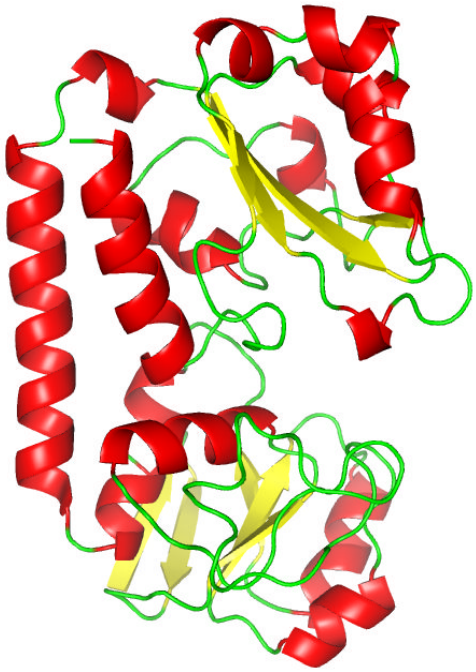


Figure 1-6: Crystal structure of FeuA, a ferric-bacillibactin lipoprotein receptor of *Bacillus subtilis* (PDB code: 2PHZ).

1.4.8 : Iron release from ferric siderophores

The mechanism of iron release from the ferric-siderophore complex is likely to involve reduction for siderophores such as hydroxamate and α-hydroxy carboxylate siderophores[9], as Fe(II) has little affinity for these siderophore ligands. This reduction is predicted to be carried out by flavoproteins, because the reduction of the Fe-siderophore complex at physiological pH is within the range of reducing agents such as flavin adenine dinucleotide (FAD)[9]. Although many enzymes with ferri-siderophore reductase activity have been identified, only a few have significant specificity for siderophore substrates, suggesting that reductive iron release has not led to the development of highly specialized enzymes[9].

A family of potential reductases have been identified which may have significant specificity for certain ferric-siderophores: these are known as siderophore interacting proteins (SIPs)[9]. One example is ViuB of *Vibrio cholerae*, which is essential for the utilization of the catecholate siderophore ferric vibriobactin[39], and a crystal structure of a related SIP[40] from *Shewanella putrefaciens* has FAD bound in a cleft (Figure 1-7). This may be the site of Fe-putrebactin reduction. An unrelated iron-sulfur

protein in *E. coli* (FhuF) has been shown to have specificity to ferric hydroxamates such as ferrichrome[41]. It is speculated that the role of the iron-sulfur cluster is to mediate electron transfer to the Fe-siderophore to reduce it.

Ferric-enterobactin and ferric-bacillibactin, in contrast, have very low redox potentials and much lower dissociation constants from Fe^{3+} (ferrioxamine E has a dissociation constant for ferric iron of 10^{-32} mol L^{-1} compared to 10^{-52} mol L^{-1} for ferric enterobactin[11]). Thus, enzyme-catalyzed hydrolysis of the backbone is the proposed iron release mechanism in *E. coli* and *Bacillus subtilis* for ferric-enterobactin and ferric-bacillibactin respectively[34,42].

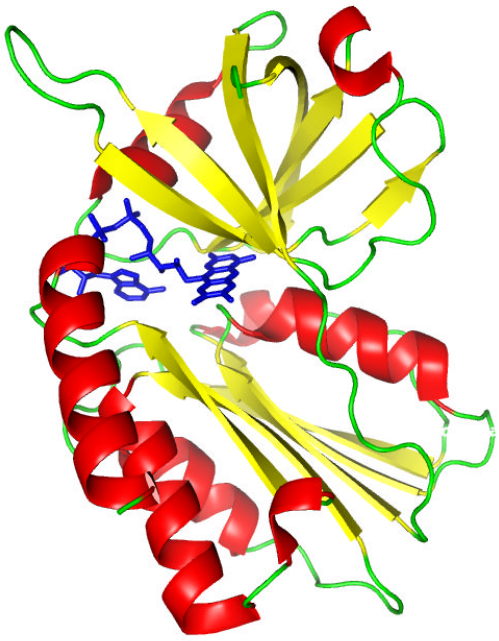


Figure 1-7: X-ray crystallographic structure of the putative SIP from *Shewanella putrefaciens.* The FAD group bound in a cleft of the protein is highlighted in blue (PDB code: 2GPJ).

1.5 : The importance of understanding siderophore-mediated iron uptake in bacteria

Siderophores have recently attracted interest as potential drug delivery agents, due to the high specificity and efficiency of their uptake systems[43]. Albomycin (Figure 1-8), a siderophore-antibiotic conjugate produced by the bacterium *Streptomyces griseus*, is effective against *E. coli* and *S. aureus* because its uptake uses the same siderophore transport system as the fungal siderophore ferrichrome[23,44].

Understanding the specificity of these transport systems could lead to the design of new "Trojan horse" antibiotics which masquerade as siderophores[43].

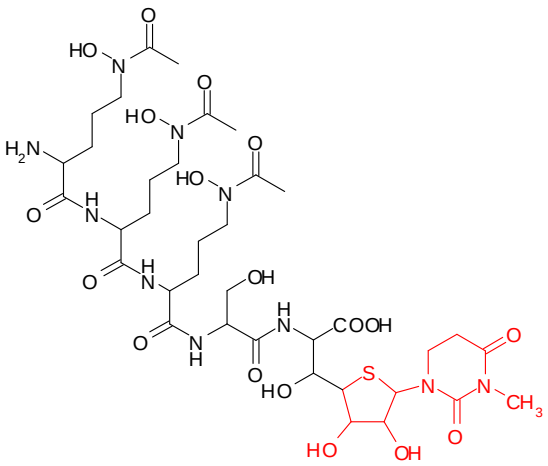


Figure 1-8: Albomycin δ_1. The antibiotic portion of the molecule is highlighted in red.

The production of siderophores is also a critical factor for pathogenic organisms to establish infections in the body[9]. For example, the Gram-positive anthrax pathogen *Bacillus anthracis* produces two siderophores, bacillibactin and petrobactin, to scavenge ferric iron. Whilst bacillibactin has been shown to bind to the immune system protein siderocalin[45] (see Figure 1-9), rendering it ineffective as a means of scavenging iron from the host, petrobactin was proposed to evade siderocalin and has been shown to be important for virulence in mice[46].

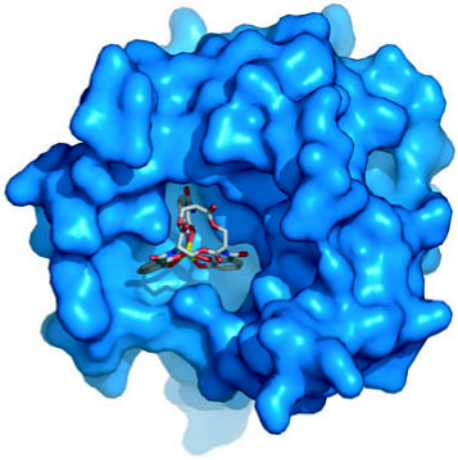


Figure 1-9: Crystal structure of ferri-enterobactin bound to the immune system protein siderocalin.

Enterobactin is an analogue to the *Bacillus anthracis* siderophore bacillibactin. Figure from [19]

The inhibition of the production of stealth siderophores such as petrobactin could lead to the development of novel antibiotics that target microbial iron acquisition pathways. Genes involved in petrobactin production have been identified, and currently its biosynthetic pathway is being elucidated by gene knockout experiments[47] and biochemical experiments with purified recombinant proteins[48]. Another approach to inhibit siderophore-mediated iron acquisition is to design inhibitors to block the uptake of ferric-petrobactin. The uptake pathway for ferric-petrobactin has been recently elucidated through binding studies involving the purified recombinant siderophore-binding lipoprotein receptors of the closely-related pathogen *Bacillus cereus*[49]. Two lipoprotein receptors have been shown to bind ferric-petrobactin and its photoproducts. Other lipoprotein receptors have been found to bind the xenosiderophores ferric enterobactin, ferric-schizokinen, ferrioxamine B and ferrichrome[49].

1.6 : Iron metabolism in the model Gram-positive organism *Streptomyces coelicolor* A3(2)

1.6.1 : Introduction

Streptomyces coelicolor A3(2) (Figure 1-10) belong to the Streptomycetes – a group of soil-dwelling filamentous Gram-positive bacteria that have a complex life cycle and are responsible for producing many antibiotics as secondary metabolites. Streptomycetes belong to the Actinobacteria, an order of GC-rich Gram-positive bacteria which include the pathogens *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae.* *S. coelicolor* is not pathogenic however (it is found in many soils and produces geosmins[50], which are responsible for the odour of soil). Its biology is relatively well characterized, and its genome was published in 2002[51].

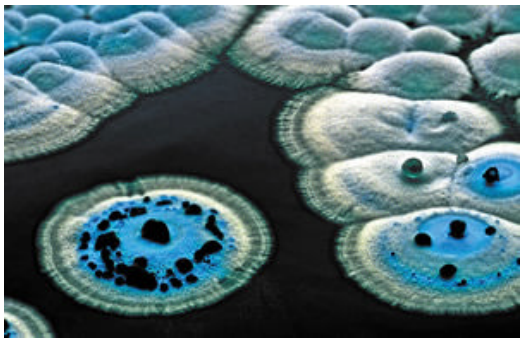


Figure 1-10: Colonies of *Streptomyces coelicolor* A3(2) growing on solid media.

The blue spots are water droplets containing the pigmented antibiotic actinorhodin. Photograph from the John Innes Centre, Norwich, UK.

Like some pathogenic microorganisms, *S. coelicolor* produces more than one siderophore (*M. tuberculosis* produces both mycobactin and carboxymycobactin[21]). In particular, *S. coelicolor* produces three siderophores of the hydroxamate type (Figure 1-11): desferrioxamine B, desferrioxamine E and coelichelin[52,53] under iron-limiting conditions. Many streptomycetes produce desferrioxamines and many more types of bacteria can utilize their ferric complexes as iron sources. This led to the proposal that coelichelin may be produced by the relatively slow-growing *S. coelicolor* (and the related *S. ambofaciens* ATCC28377) to guarantee a ferric iron supply in the competitive soil environment[18].

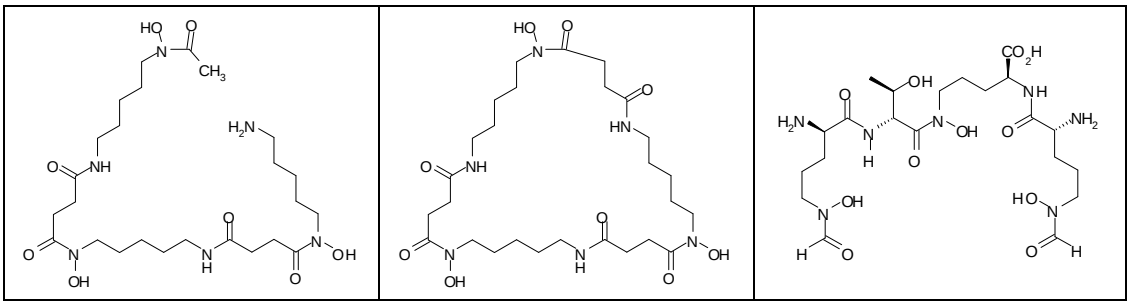


Figure 1-11: Siderophores produced by *Streptomyces coelicolor*.

Left: Desferrioxamine B, marketed as its mesylate salt Desferal®. Middle: Desferrioxamine E; Right: Coelichelin

1.6.2 : Biosynthesis and uptake of siderophores in *Streptomyces coelicolor* A3(2)

1.6.2.1 : **The desferrioxamines**

The *des* cluster of genes in *S. coelicolor* (Figure 1-12) is responsible for the production of both desferrioxamines E and B[54]. A role of each of the genes *desABCD* in their biosynthesis have been proposed[54]. DesD belongs to the group of NRPS-independent siderophore (NIS) synthetases involved in siderophore production[15,55]. Immediately upstream of *desA* is a DmdR (divalent metal-dependent regulatory protein) binding site and its function has been experimentally elucidated, which suggest that the expression of the genes is controlled by intracellular ferrous iron levels[56].

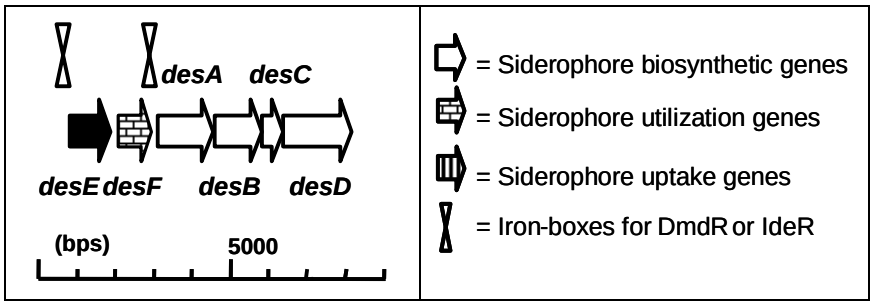


Figure 1-12: The organisation of the *des* cluster of *S. coelicolor*

Despite the structural similarities between the desferrioxamines B and E, they have distinct properties – desferrioxamine E has a 100-fold greater affinity for Fe(III) than desferrioxamine B[11], and the latter is the only one that is used clinically to treat iron overload[12].

Whilst *desABCD* is involved in the biosynthesis of desferrioxamines, there were also two genes upstream of *desA*, namely *desE* and *desF*, which also appear to be transcriptonally controlled by Fe(II) levels inside the cell based on the presence of a DmdR binding site upstream of *desE*. These genes were predicted to encode a lipoprotein receptor (siderophore-binding protein) component of an ABC transporter and a SIP likely to be involved in the reduction of ferric siderophores, respectively (Table 1-2)[53]. The permease and ATPase domains of the DesE-associated ABC transporter are not encoded by genes in the gene cluster, suggesting DesE would have to interact with other (unidentified) permeases to transport ferri-siderophores.

Table 1-2: Proposed functions of proteins encoded by the *des* cluster[53].

Protein	Homologue (% identity)	Proposed function
DesA	Ddc *A. baumannii* (37%)	L-Lysine decarboxylase
DesB	AlcA *B. bronchiseptica* (47%)	1,5-Diaminopentane monooxygenase
DesC	AlcB *B. bronchiseptica* (46%)	Acyl-CoA dependent acyl transferase
DesD	AlcC *B. bronchiseptica* (51%)	Type C siderophore synthetase
DesE	FhuD *B. subtiltis* (24%)	Ferric-siderophore lipoprotein receptor
DesF	ViuB *V. cholerae* (30%)	Siderophore-interacting hydrolase

1.6.2.2 : **Coelichelin**

Coelichelin production is directed by the *cch* cluster of genes[52], and is also found in *S. ambofaciens* (Figure 1-13). The largest gene by far in this cluster is the *cchH* gene, encoding a trimodular NRPS.

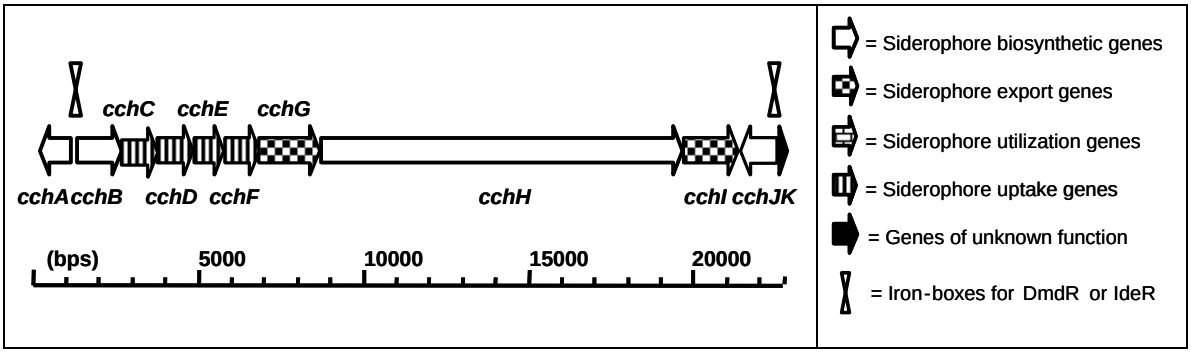


Figure 1-13: The organisation of the coelichelin biosynthetic cluster in *S. coelicolor*.

The unusual feature of the CchH NRPS are that it lacks a C-terminal thioesterase domain which is usually involved in the release of the assembled peptide from the synthetase, and that it produces a tetrapeptide rather than the predicted tripeptide[57]. A novel mechanism for coelichelin biosynthesis was formulated, where the non-proteinogenic amino acid *N*-formylhydroxyornithine was loaded twice onto the same module of the NPRS to make the tetrapeptide, and the enzyme encoded by the *cchJ* gene (similar to *E. coli* enterobactin esterase – see Table 1-3) was implicated in releasing the tetrapeptide from the enzyme[52]. The genes *cchCDEF*, which encode an ABC transporter, are flanked by DmdR iron boxes. *cchF* encodes a lipoprotein receptor, *cchC* and *cchD* encodes the permeases, and *cchE* encodes an ATPase.

1.6.3 : The effect of mutations in the *des* and *cch* gene clusters on siderophore uptake

To investigate the effect of deletion of the *des* and *cch* gene clusters in *S. coelicolor* on siderophore uptake[53], the gene clusters were separately replaced by apramycin and viomycin resistance cassettes respectively using the PCR-targeting REDIRECT technology[50]. Siderophore non-producing mutants that had both genes *desD* and *cchH* replaced were also made. The cluster deletion mutants lacked the DesE and CchF lipoprotein receptors, while the *desD* and *cchH* mutants still encoded these receptors. The various mutants were grown on iron-deficient colloidal silica plates (which lacked xenosiderophores), and the growth halos around filter paper discs supplemented with different siderophores were observed.

Table 1-3: Proposed functions of proteins encoded by the *cch* cluster[53]

Protein	Homologue (% identity)	Proposed function
CchA	FxbA *M. smegmatis* (49%)	Formyl transferase
CchB	PsbA *Pseudomonas* sp. B10 (39%)	L-Ornithine-N5-monooxygenase
CchC	FepD *Yersinia* spp. (39%)	Permease component of ABC importer
CchD	FepG *Yersinia* spp. (35%)	Permease component of ABC importer
CchE	FepC *Y. enterocolitica* (54%)	ATPase component of ABC importer
CchF	FhuD *B. subtiltis* (26%)	Ferric-siderophore lipoprotein receptor
CchG	ExiT *M. smegmatis* (50%)	ATPase/permease of exporter
CchH	FxbC *M. smegmatis* (42%)	Coelichelin NRPS
CchI	StrV *S. glaucescens* (33%)	ATPase/permease of exporter
CchJ	Fes *E. coli* (30%)	Thioesterase
CchK	MbtH *M .tuberculosis* (70%)	Essential for biosynthesis[58]

The *des* cluster mutant was unable to grow well when supplemented with ferrioxamine E and ferrichrome, whilst the *cch* cluster mutant could not grow when supplemented with ferri-coelichelin. Surprisingly, all of the mutants, including ones

where both *des* and *cch* clusters were deleted, were able to grow well with ferrioxamine B, coprogen and aerobactin supplementation. This suggested the presence of another transport system, independent of the *des* and *cch* clusters, which were involved in ferrioxamine B uptake and utilization.

1.6.4 : Identification of another ABC transporter involved in ferrioxamine B transport

Bunet *et al.*[59] identified another gene cluster *cdtABC* that could be involved in siderophore transport (Figure 1-14). A putative DmdR binding site was identified upstream of *cdtC*[53], and the *cdtB* gene encoding a protein similar to siderophore binding proteins (Table 1-4).

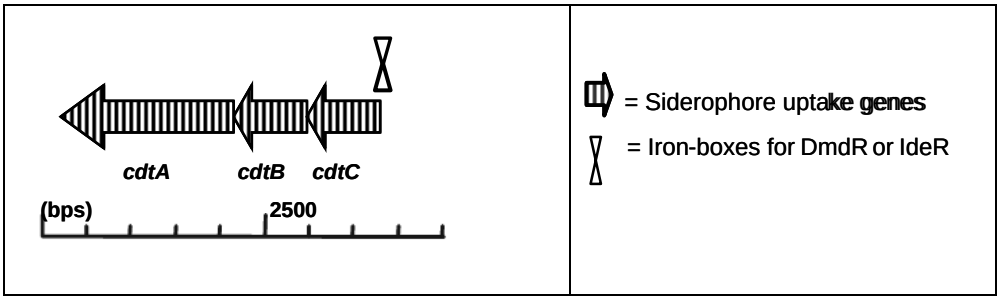


Figure 1-14: The organisation of the *cdtABC* cluster in *S. coelicolor* which plays a role in salmycin uptake and possibly ferrioxamine B uptake.

To determine whether this gene cluster had a role in siderophore transport, the *cdtABC* cluster was replaced using the REDIRECT technology[60] with an apramycin resistance cassette, and the resulting mutant was grown in the presence of ferri-salmycin[44], a siderophore-antibiotic conjugate analogous to ferrioxamine B (Figure 1-15). It was found that the mutant was not sensitive to ferri-salmycin, whilst the wild-type *S. coelicolor* was, and that this was not due to the presence of the apramycin-resistance gene used to replace *cdtABC.* Competition experiments with ferrioxamine B found that this siderophore reduced the sensitivity of the wild-type *S. coelicolor* to salmycin on solid media, although uptake experiments with radioactive ^{55}Fe-ferrioxamine B suggested there was no difference in uptake of this siderophore between the wild-type and the mutant.

Table 1-4: Proposed functions of proteins encoded by the *cdt* cluster of genes[59].

Protein	Homologue (% identity)	Proposed function
CdtA	*E. coli* FepG (33%) *E. coli* FepD (33%)	Permease dimer/component of ABC importer
CdtB	*E. coli* FepB (24%)	Ferric-siderophore lipoprotein receptor
CdtC	*E. coli* FepC (41%)	ATPase component of ABC importer

These results suggested that the antibiotic salmycin is taken up by *S. coelicolor* via the *cdtABC* transporter system. This system can also take up ferrioxamine B, although ferrioxamine B could be taken up via an alternative route – possibly via DesE, which is involved in ferrioxamine E uptake.

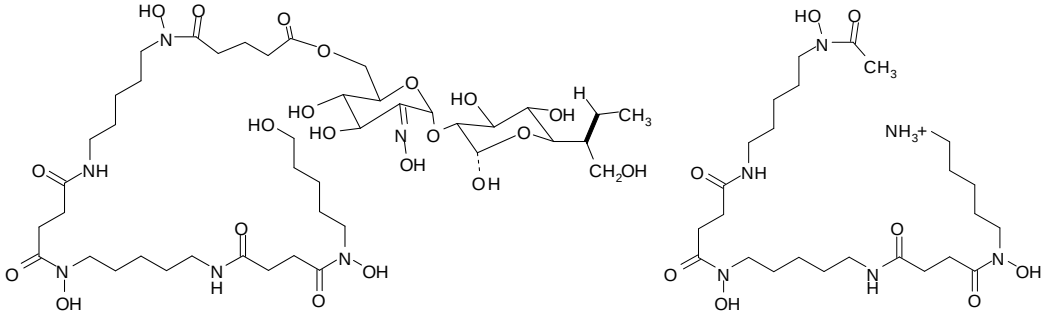


Figure 1-15: The structure of salmycin A compared to desferrioxamine B

1.6.5 : Other genes which may be involved in siderophore uptake

Gene clusters were also found that had similarity to the *E. coli* enterobactin ABC transporter system FepBCDG. SCO1785-1787 had similarities to the ATPase FepC, permease FepG and the permease FepD, whilst SCO2272 had similarities to the SBP FepB, FepD and FepC respectively. Also the genes SCO0996 and SCO7218 had similarities to SBPs, and were adjacent to ABC transporter system genes.

1.6.6 : The proposed roles of the DesE, CchF and CdtB lipoprotein receptors in siderophore uptake

From the gene knock-out experiments, the following model was proposed[53]: After excretion of the siderophores and formation of the corresponding ferric complexes,

the lipoprotein receptor DesE was proposed to bind to ferrioxamine E, and possibly ferrioxamine B, but not to ferri-coelichelin. The transport of the ferri-siderophore complex is mediated by as yet unknown ABC transporter permease and ATPase. The ABC transport system CchCDEF, with receptor CchF, transports only ferri-coelichelin, whilst the CdtABC system, containing the receptor CdtB, could transport ferrioxamine B, but not ferrioxamine E or ferri-coelichelin. Also, the hydroxamate siderophore ferrichrome was proposed to bind only to DesE, but not CchF and CdtB. The proposed biosynthesis and uptake of cognate siderophores of *S. coelicolor* is illustrated in Figure 1-16.

1.6.7 : Methods for investigating the role of the lipoprotein receptors

To elucidate the routes of ferri-siderophore transport in *S. coelicolor* A3(2), one has to study the lipoprotein receptors which govern ferri-siderophore entry. To test the interaction of siderophore-binding proteins with ferri-siderophores, the main approach used by others have been to recombinantly express and purify the proteins in the bacterial host *E. coli*, and then to monitor the decrease of intrinsic fluorescence of the proteins on titration with small amounts of ferri-siderophores[30,38,49,61]. The sensitivity of the fluorescence technique has been used to calculate dissociation constants in the nanomolar range, and required only small amounts of protein and ferri-siderophores.

If there is binding of a siderophore-binding protein to a ferri-siderophore, it is likely that the ferri-siderophore is taken up into the cell via its corresponding ABC transporter components, whilst if there is little or no binding, one can deduce that the siderophore-binding protein is not involved in the uptake of that ferri-siderophore.

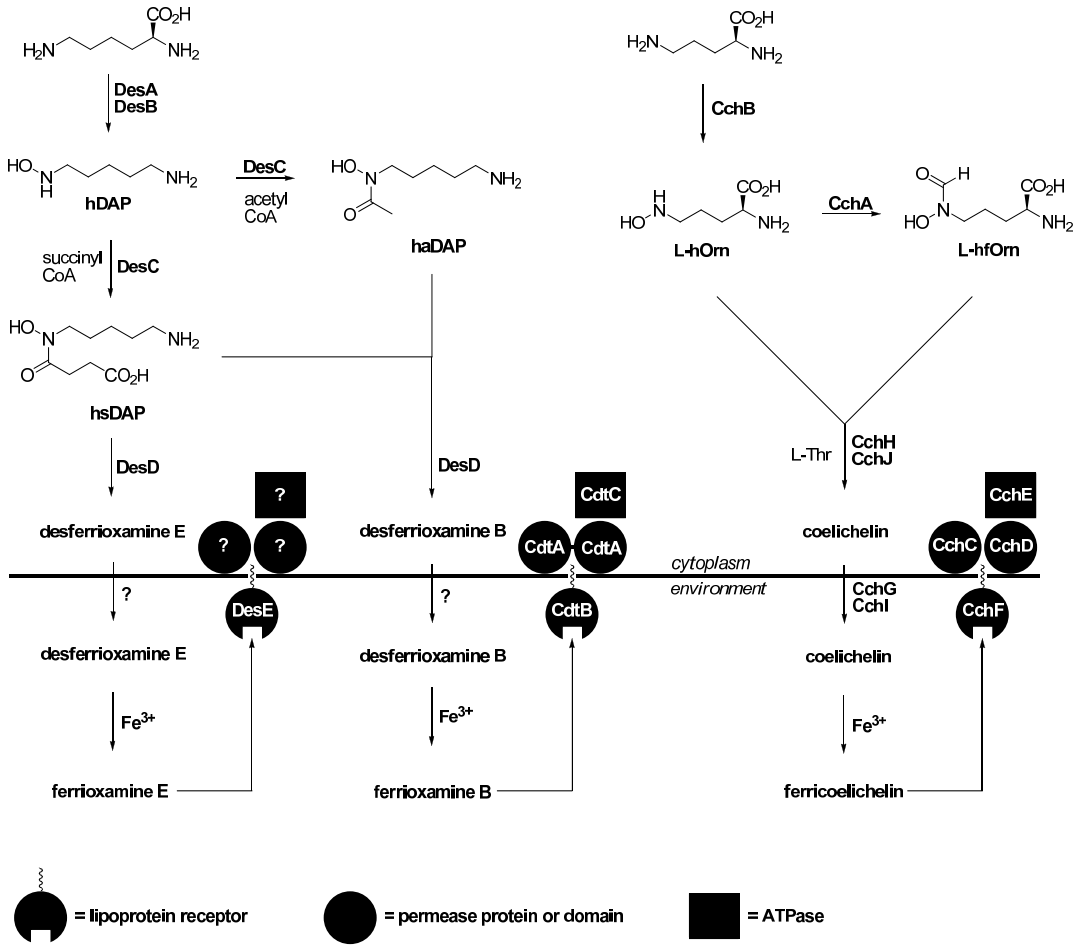


Figure 1-16: Model of the biosynthesis, export and uptake of coelichelin, desferrioxamine E and desferrioxamine B by *Streptomyces coelicolor* A3(2).

Abbreviations for intermediates in coelichelin and desferrioxamine biosynthesis are as follows: hDAP: N-hydroxy-1,5-diaminopentane; haDAP: N-hydroxy-N-acetyl-1,5-diaminopentane; hsDAP: N-hydroxy-N-succinyl-1,5-diaminopentane; L-hOrn; L-N5-hydroxyornithine; L-N5-hydroxy-N5-formylornithine. An arrow indicating the possibility that DesE is also involved in ferrioxamine B binding is not shown. Figure from [53].

1.6.8 : Properties of the three putative lipoprotein receptors DesE, CchF and CdtB

The protein sequence of DesE, CchF and CdtB proteins share similarities to other siderophore binding proteins from Gram-positive and Gram-negative bacteria (Figure 1-17). A multiple alignment of DesE, CchF and CdtB proteins with *E. coli* FhuD and *B. subtilis* FeuA showed several regions of great similarity, but these correspond to structural elements (α-helices and β-sheets) that are important as the

scaffold of the FhuD and FeuA proteins. These residues do not participate in ligand binding. Also, there were several glutamate and aspartate residues not in α-helices or β-sheets that were conserved - in FhuD and FeuA these correspond to the permease-binding face of these proteins, which may be important for interactions[26].

```
DesE    -MSHASATHP TRRGILAAGG ALGLGAVLAA CGDGDGKSDG AGDGSGGAAK SGPWSFKDDR
CchF    MLLRTTRTKP WRRLAAALSA AALGVGLLAG CGSDSDDPAD EAGGGTPAAA GAFPVTVEHA
CdtB    ---------- MRRLLLTAAA TTAAALTLAA CGTTEPAADK AEKKASEAI- ----TLKDGK
FeuA    ---------- -MKKISLTLL ILLLALTAAA CGSKNESTAS KASGTASEK- -----KKIEY
FhuD    ---------- MSGLPLISRR RLLTAMALSP LLWQMNTAHA AAIDP----- ----------

DesE    GTTVKLDKVP ANIVAFTG-V AAALFDYGVE VKGVFGPTTT KDGKPD---- ----VQAGDL
CchF    FGTTKIDKAP ERVVSVGYTD DQTVLAFGIK PVGMVDQYPN PAGQSPDINT QWPWVKDKWG
CdtB    GTEVKLDGPA TKVVATEWNV VESLVSLGVD PVGVADVKGY KTWDSA---- ----VPLKNE
FeuA    LDKTYEVTVP TDKIAITG-S VESMEDAKLL DVHPQGAISF SGKFPD---- ----MFKDIT
FhuD    ---------- NRIVALEWLP VELLLALGIV PYGVADTINY RLWVSE---- ----PPLPDS

DesE    DVDKVTVLGN EWGKLNVEKY ASLAPEVLIT TTFDTAGTLW SVPEESKDKV AKLAPSVAIS
CchF    DTKPEVVMKN GDTGPNFEKI AALRPDLIVA VYSEID---- ---QAAYDKL SKIAPTVGRT
CdtB    PKD----IG- TRGEPSMDTV ASLAPDLIVA TTDLAP---- ----AAVKQL REVAPVIEVK
FeuA    DKAEP--TG- EKMEPNIEKI LEMKPDVILA STKFPE---- ----KTLQKI STAGTTIPVS
FhuD    VID----VG- LRTEPNLELL TEMKPSFMVW SAGYGP---- ----SPE-ML ARIAPGRGFN

DesE    VFDR--QLTQ PLQRMWELAE SLGADMKAKK VTDAKAAFDK AAARLRAAAK AKPEIRVLAG
CchF    KGEKEPFSAP WQDNALHIAK ALGKAEEGEK MVADIQGKLD AAKQAHPEFA DKTAVVLSWY
CdtB    SADGTGQIDR MLENVDLIAE ATGTTDRAKS LREGFETKVA EGKKALADAG MAGKDIAFAD
FhuD    FSDGKQPLAM ARKSLTEMAD LLNLQSAAET HLAQYEDFIR SMKPRFVKRG AR--PLLLTT
FeuA    HISS-----N WKENMMLLAQ LTGKEKKAKK IIADYEQDLK EIKTKINDKA KDSKALVIRI

DesE    ----SASPDL FYVSGTNLSV DLEYFKALGV NFVEPSEDAK KATGGWFESL SWENVDKYPA
CchF    K------DSV APFTSTDVRG --RLVTGIGF KYETEIDKVA GGDFYTTLSP ERVDLVDVDR
CdtB    GYVASNQVSI RPYTATSLIG EVNEAVGLKN AWTVKGDEAY GLGATDVEGL TKLPEDTQFA
FeuA    R------QGN IYIYPEQVYF NSTLYGDLGL KAPNEVKAAK AQELSSLEKL SEMNPDHIFV
FhuD    ---LIDPRHM LVFGPNSLFQ EILDEYGIPN AWQGETN-FW GSTAVSIDRL AAYKDVDVLC

DesE    DVIIMDDRAS TIQPADITEG -TWKQLPAVK AGQVIARSPE PILSYDKCTP LLDNLAE--A
CchF    VFVINDKAD- ---QDALKKF ELFTNLDAVK NGKVSYLLDS EGPAVGAAIS QGTLLSMPYA
CdtB    YIGNDDDPSA TPFTGELAKN PVWKSLPFVK AG-DVHRLDD GIWMFGGPGS MEAYIDA--V
FeuA    QFSDDENADK PDALKDLEKN PIWKSLKAVK EDHVYVNSVD PLAQGGTAWS KVRFLKA--A
FhuD    FDHDNSKD-- ---MDALMAT PLWQAMPFVR AG-RFQRVP- AVWFYGATLS AMHFVRV--L

DesE    IENAKKVG-
CchF    VDELVKSAG
CdtB    VGALTK---
FeuA    AEKLTQN--
FhuD    DNAIGGKA-
```

Figure 1-17: Multiple alignments of DesE, CchF and CdtB with the siderophore-binding proteins *E. coli* FhuD and *B. subtilis* FeuA. Key: Magenta: conserved residues throughout the five proteins. Green: conserved residues from three or more proteins. Cyan: Similar residues to the green residues. The alignment was performed by ClustalX using Gonnet scoring.

The DesE, CchF and CdtB proteins were found in the *S. coelicolor* membrane-associated proteome[62] by analysis of membrane protein extracts in liquid culture using two-dimensional gel electrophoresis (Figure 1-18), and matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) peptide mass fingerprinting of the resulting protein spots. All three proteins ran on the gel at similar molecular masses and isoelectric points around pI 4 – 5, as predicted by their protein sequences

(Table 1-5). This data is consistent with the proposed role of these proteins as membrane-associated receptors.

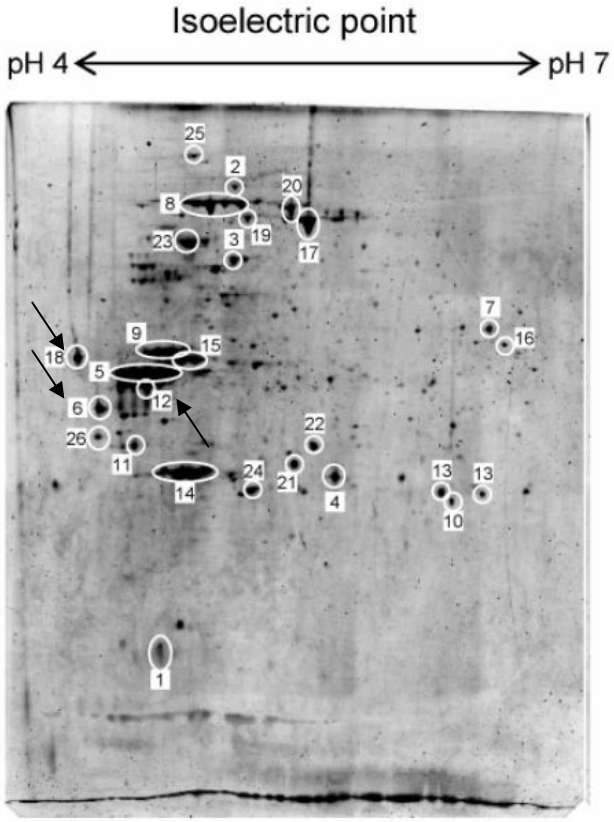


Figure 1-18: Two dimensional gel electrophoresis of the membrane-associated proteome of *S. coelicolor.*

The spots 6, 12 and 18 (indicated by arrows) were identified by peptide mass fingerprinting as the proteins CchF, DesE and CdtB respectively. Figure from [62].

Also, all three proteins have been identified as putative substrates of the twin-arginine translocase (TAT) protein exporter of *S. coelicolor*[63]. This suggests that these proteins are produced and folded in the cytoplasm of *S. coelicolor*, and exported in a fully folded form to the outer membrane. TAT substrates are identified by their hydrophobic N-terminal signal sequences (Table 1-6) with the distinctive RR (twin arginine motif)[64]. To test whether DesE, CchF and CdtB were TAT substrates, the membrane-associated proteome of *S. coelicolor ΔtatC* (defective in TAT substrate export) was compared to the wild-type strain. Only DesE was clearly absent from the *ΔtatC* membrane proteome, and its TAT signal sequence was able to direct the export of the agarose-consuming enzyme agarase on solid media when a chimeric gene was integrated into the wild-type *S. coelicolor* chromosome[63].

Table 1-5: The molecular masses and pIs of the three lipoprotein receptors predicted from their protein sequences (prior to N- terminal cleavage and lipidation)

Protein	Molecular weight (Da)	pI
DesE	36641.8	4.94
CchF	37374.6	4.73
CdtB	34093.9	4.44

Analysis of the sequences of DesE, CchF and CdtB indicate the presence of the prokaryote lipidation site PS51257[65] (also known as a lipobox). This suggests the proteins are likely to be lipidated at the N-terminal cysteine after translocation to the outer membrane by the Lgt protein of *S. coelicolor* and N-terminal cleavage catalysed by the Lsp protein[65](which may also acylate the free amino group of the cysteine) (Figure 1-19).

Table 1-6: The first 35 residues of the *S. coelicolor* DesE, CchF and CdtB proteins.

The twin-arginine translocase (TAT) consensus sequence R-R-x-Φ-Φ (where x is any amino acid and Φ is a hydrophobic residue) is highlighted in cyan, while the PS51257 Prosite prokaryotic membrane lipid attachment site (originally called PS0013) is highlighted in green. The lipid is attached to the cysteine residue at the C-terminus of the highlighted sequence.

Protein	First 35 residues of the protein sequence
DesE	MSHASATHPTRRGILAAGGALGLGAVLAACGDGDG
CchF	MLLRTTRTKPWRRLAAALSAAALGVGLLAGCGSDS
CdtB	MRRLLLTAAATTAAALTLAACGTTEPAADKAEKKA

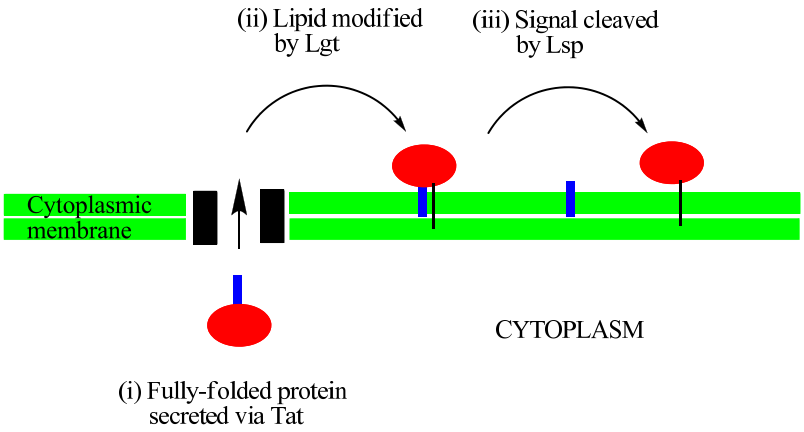


Figure 1-19: The proposed mechanism of lipidation of the lipoprotein receptors in Gram-positive bacteria.

(i) The protein (red) with hydrophobic signal sequence (blue) is secreted by the TAT pathway to the cytoplasmic membrane; (ii) The Lgt protein then lipidates the N-terminal cysteine of the protein at the end of the signal peptide (black line); (iii) the signal peptide is cleaved by Lsp, leaving the mature protein (red) with an N-terminal lipid (black).

1.7 : Structural aspects of SBP-siderophore binding

In *S. coelicolor*, the DesE receptor was predicted to bind ferrioxamines B and E, and ferrichrome, whilst CdtB was predicted to bind ferrioxamine B only. CchF was predicted to bind only ferri-coelichelin. It is difficult to predict which siderophores a receptor may bind to from its sequence alone; one can only infer the protein may bind to siderophores. For example, CdtB was annotated as a ferrichrome-binding protein due to its similarity to *E. coli* FhuD, although the available experimental evidence suggests otherwise. What features do siderophore-binding proteins have to make them so specific for various ferri-siderophores?

The recognition of ferric siderophores by OMRs, SBPs and lipoprotein receptors is highly dependent on the coordination environment of the ferric iron[11]. *E. coli* can take up the *tris*-hydroxamate siderophores ferrichrome, ferrichrysin and ferricrocin (whose peptide backbones differ, but the coordinate environment of ferric iron is similar), but cannot use ferrichrome A. This is due to the periphery groups attached to the hydroxamate groups in ferrichrome A (Figure 1-20), which affects its binding to the *E. coli* OMR FhuA[66].

The chirality of ferri-siderophores around the iron centre, in particular, the handedness of the propeller when the Fe-siderophore is viewed from the C_3 axis (Figure 1-21), has also been marked out as an important factor for Fe-siderophore uptake[67]. Inhibition of ferric coprogen uptake in the fungus *Neurospora crassa* was greater with Δ-*trans* isomers of chromic coprogen than with Δ-*cis* isomers, implying that Δ-*trans* coprogen is recognised by its receptors better than Δ-*cis* coprogen[67]. The ferric complex of coprogen crystallized as the Δ-*trans* isomer[68].

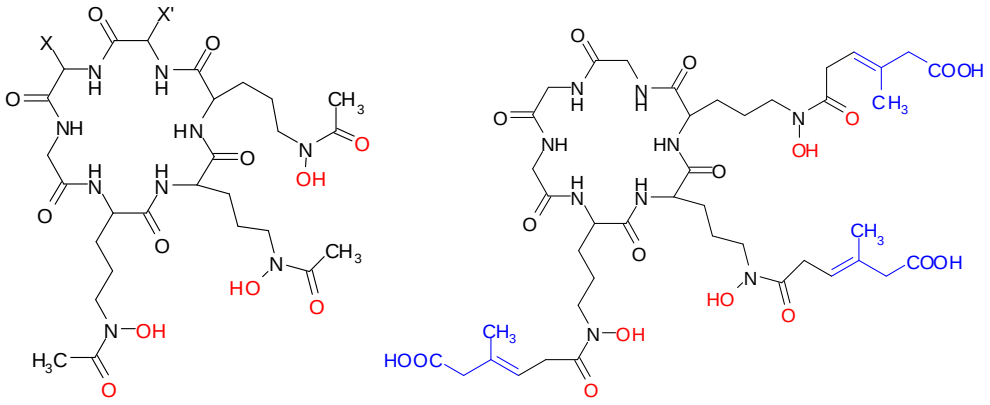


Figure 1-20: Desferri-siderophores of the ferrichrome family.

Left: Desferri-ferrichrome (X=H, X'=H), desferri-ferrichrysin (X=CH_2OH, X'=CH_2OH) and desferri-ferricrocin (X=CH_2OH, X'=H). Right: Desferri-ferrichrome A. The iron-ligating regions are highlighted in red. *E. coli* cannot use ferrichrome A as an iron source due to the presence of the groups highlighted in blue.

On the other hand, the crystal structure of *E. coli* FhuD with the siderophores ferrioxamine B, ferric-coprogen and ferri-albomycin bound showcased various coordination environments of ferric hydroxamates on binding to the SBP, with the complexes exhibiting either the Δ or Λ configuration. This may be due to the relaxed substrate specifity of this protein, by virtue of having to shuttle siderophores collected by both the FhuA (ferrichrome)[69] and FhuE (coprogen) OMRs[70].

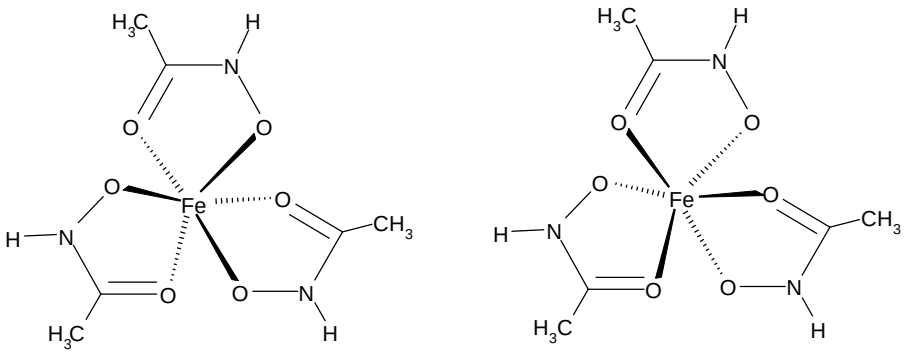


Figure 1-21: The Λ and Δ configurations of Fe(acetohydroxamic acid)$_3$, a simple ferri-hydroxamate

To attempt to understand siderophore-SBP binding fully, one must be able to produce a model of the SBP bound to their putative substrates, usually via X-ray diffraction of a crystal of the complex. This is a time-consuming exercise, as many proteins do not crystallize readily. Once a crystal structure is found, one can analyse the hydrogen bonds and van der Waals' contacts of the siderophore to the protein.

1.8 : Molecular modelling of ferric-tris-hydroxamates to predict the strength of siderophore-SBP binding

Once the structure of a receptor/ferri-siderophore complex has been determined, one may want to predict the affinity of the protein for other ferri-siderophores by molecular modelling. As the coordination environment of the iron is crucial for its recognition by the SBP, the molecular model must be able to reproduce the bond lengths, angles, torsion angles etc. of ferri-siderophore complexes. The problem with many molecular modelling programs is that they cannot model the high coordination numbers of transition metal complexes, and the hexacoordinate Fe(III) siderophore complexes are assumed to be perfectly octahedral, which is not the case (Figure 1-22).

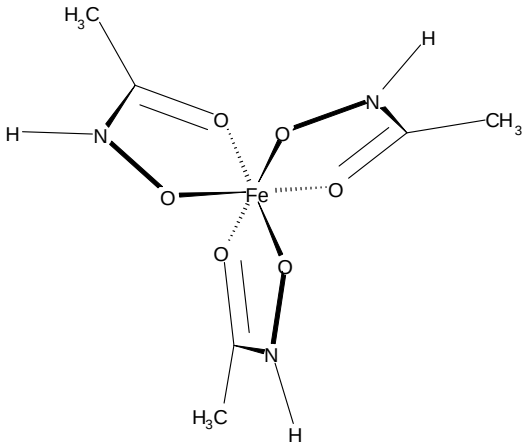


Figure 1-22: A model of the simple ferric-*tris*-hydroxamate, Fe(acetohydroxamate)$_3$.

The arrangement of the oxygen atoms around the Fe centre in the X-ray crystal structure[71] is not octahedral (O_h symmetry), but very close to C_3 symmetry. This complex is representative of ferric hydroxamate ligands bound to Fe(III).

Also, a model should be able to accurately reflect the lability of the coordination environment of iron with its siderophore ligands, due to the low ligand field strength. Ferrioxamine B formed a racemic mixture of Λ-cis and Δ-cis (*fac*) isomers when crystallized on its own[72], but bound to *E. coli* FhuD it adopts the Λ-C-*trans-cis* configuration[28]. To investigate protein-siderophore binding computationally, the transition of siderophores from *fac* isomers to other isomers has to be modelled effectively. This involves the careful modelling of the energy of trigonal distortion of the complexes, because this is the most probable way that isomers interconvert.

The software package DOMMIMOE[73], developed at the University of Warwick, is an incorporation of the ligand field theory of transition metal complexes into the Molecular Operating Environment (MOE)[74], a commercial molecular mechanics package. It is an attractive tool for developing a suitable parameterization for ferric-*tris*-hydroxamate siderophores.

1.9 : Aims of the project

The main aim of this project was to characterize DesE, CchF and CdtB, the three lipoprotein receptors of *Streptomyces coelicolor* A3(2), and assess their binding with the cognate siderophores ferrioxamine B, ferrioxamine E and ferri-coelichelin, as well as the xenosiderophores ferrichrome, albomycin and ferric enterobactin. The

siderophore ferri-coelichelin is not commercially available. Thus it was planned to purify it by RP-HPLC from large-scale cultures of *S. coelicolor* grown in iron deficient medium. It was aimed to overproduce the DesE, CchF and CdtB proteins in *E. coli* and to purify them by affinity chromatography. The change in the intrinsic fluorescence of the proteins was to be used to monitor ferri-siderophore binding and to calculate dissociation constants. Another aim was to determine the structure of DesE protein with one of its substrates, ferrioxamine B, bound to gain further insights into the molecular basis of ferric siderophore recognition.

It was also planned to construct a molecular model which accurately reproduces the structural characteristics of ferric-*tris*-hydroxamate siderophores, in particular bond angles, lengths and trigonal twists of the complexes, using statistical optimization techniques to parameterize the program DOMMIMOE. This model could be used to predict and assess the binding of DesE to non-cognate ferric-*tris*-hydroxamate siderophores.

Chapter 2: Isolation and purification of ferri-coelichelin

2.1 : Introduction

Ferri-coelichelin was not commercially available, so it was isolated from cultures of *Streptomyces coelicolor.* Less than 500 μg of ferri-coelichelin was produced, using three RP-HPLC steps, involving the purification of the desferri- and gallium complexes of coelichelin[52]. As coelichelin is a siderophore, it is only produced in significant quantities if grown under iron-deficient conditions. A medium based on the iron-deficient medium used by Müller and Raymond[12] was used to grow *Streptomyces* species for siderophore production.

2.1.1 : Ferri-coelichelin production comparison of *S. coelicolor* M145, S. *ambofaciens* ATCC 23877 and *S. longisporus ruber*

S. coelicolor M145 and *S. ambofaciens* ATCC23877 have been shown to produce desferri-coelichelin in a previous study[53]. Both contain a similar *cch* cluster that directs its biosynthesis. *S. longisporus ruber* is responsible for producing the red antibiotic streptorubin A, which has a very similar structure to streptorubin B produced by *S. coelicolor*. It had been shown in unpublished work that *S. longisporus ruber* also produces coelichelin. All these strains were analysed for production of ferri-coelichelin after growth in iron-deficient medium for four days. The concentrated supernatants supplemented with $FeCl_3$ were analysed by LC-MS using the method of Barona-Gómez *et al.*[53]

2.1.2 : Results

The extracted ion chromatograms at m/z 619 contained a large peak at 3 – 4 minutes for all the culture supernatants (Figure 2-1). Analysis of the mass spectrum between 3 and 4 minutes revealed a large peak with m/z 619 which had the characteristic isotope pattern of an iron-containing species (Figure 2-2). This corresponded to a large peak at 3.1 minutes in the UV chromatogram at 435 nm (Figure 2-3), the

maximum absorbance wavelength for the charge transfer transition of ferric hydroxamates.

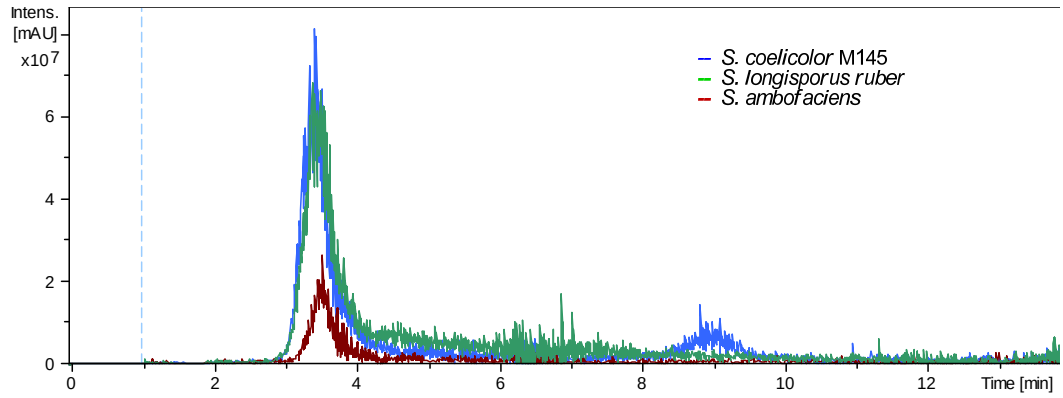


Figure 2-1: LC-MS analysis of *Streptomyces* culture supernatants: extracted ion chromatograms at m/z 619 from 0 – 14 minutes of HPLC run.

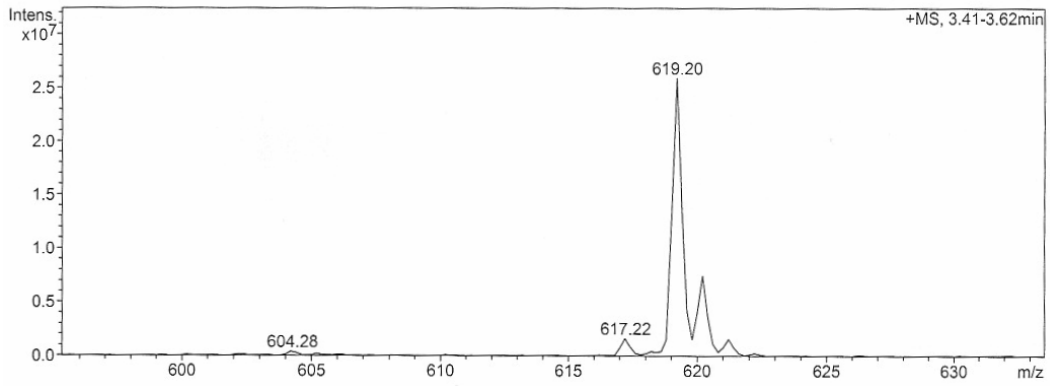


Figure 2-2: Mass spectrum from the *S. longisporus ruber* supernatant between 3.41 – 3.62 minutes from m/z 595 – 635.

The large signal at m/z 619 has a distinctive isotope pattern corresponding to an iron-containing compound, in this case ferri-coelichelin.

The retention time of 3-4 minutes was similar that reported in two previous studies[52,53], leading to the conclusion ferri-coelichelin was being detected, hence *S. longisporus ruber* did indeed produce the siderophore coelichelin (or a compound whose ferric complex has the same mass and retention time). On comparison of the UV chromatograms (Figure 2-3), the largest peak corresponding to ferri-coelichelin was observed in the *S. coelicolor* M145 supernatant; hence this strain was used to produce ferri-coelichelin for further studies.

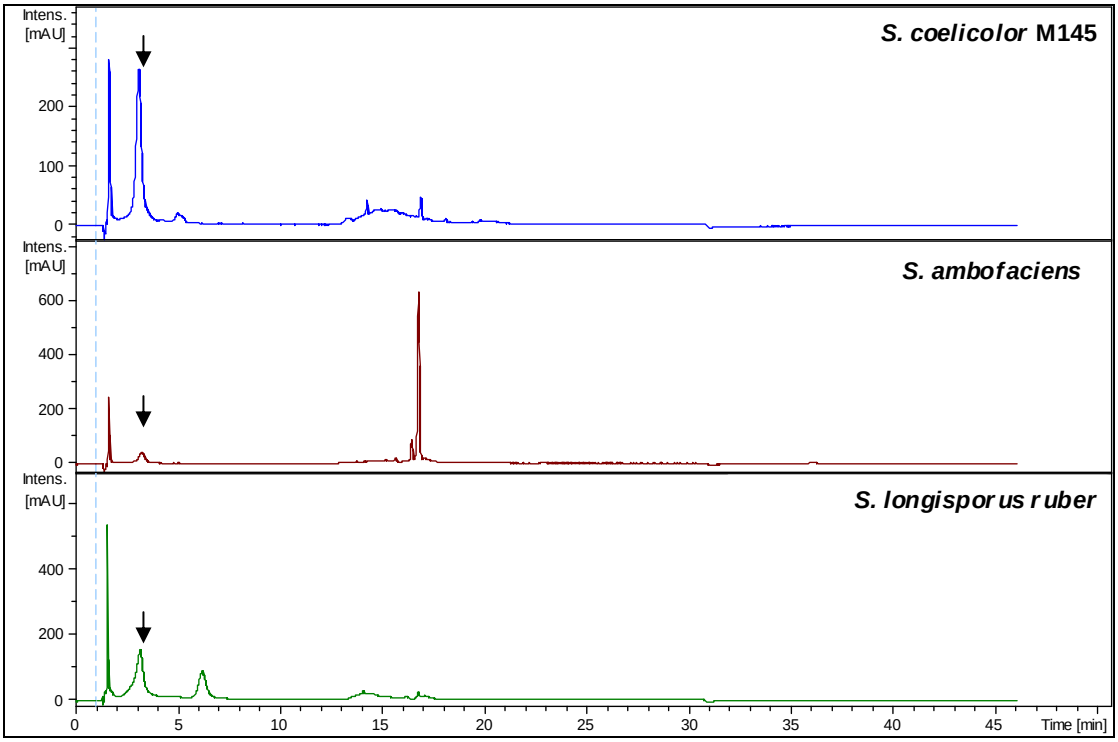


Figure 2-3: LC-MS analysis of *Streptomyces* culture supernatants: Absorbance at 435 nm

UV chromatograms (monitoring absorbance at 435 nm) from HPLC-MS analysis of the $FeCl_3$-supplemented supernatants of cultures of *S. coelicolor* M145, *S. ambofaciens* ATCC23877 and *S. longisporus ruber*. Ferri-coelichelin eluted from the HPLC column at 3.2 minutes and is indicated by the arrows.

2.2 : Large-scale purification of ferri-coelichelin

2.2.1 : Growth of *S. coelicolor* cultures

Ferri-coelichelin was produced on a large scale by growing 2 L cultures of *S. coelicolor* M145 and *S. coelicolor* W2 bacteria. The W2 strain has the *desD* gene, which is essential for desferrioxamine B and E biosynthesis, replaced by an apramycin resistance cassette, hence the only siderophore it produced was coelichelin.

2.2.2 : RP-HPLC separation of ferri-coelichelin from other siderophores

The HPLC method used to separate ferri-coelichelin from the other ferri-siderophores differed to the published procedure[52] in that a semi-preparative C-18 column was used instead of the Supelco Discovery H5-F5 column. A method was developed using a gradient of the mobile phases H_2O (0.1% TFA) and MeCN (0.1% TFA), with ferri-coelichelin eluting with mobile phases 93% H_2O and 7% MeCN. The eluate was monitored by absorption of light at 435 nm, and a yellow-orange elute was collected (Figure 2-4). The presence of ferri-coelichelin was confirmed by positive ion mode ESI-TOF-MS, Large peaks with m/z 619 and 641, indicative of the protonated and sodiated forms of ferri-coelichelin, were observed. The fractions containing ferri-coelichelin were combined and lyophilized.

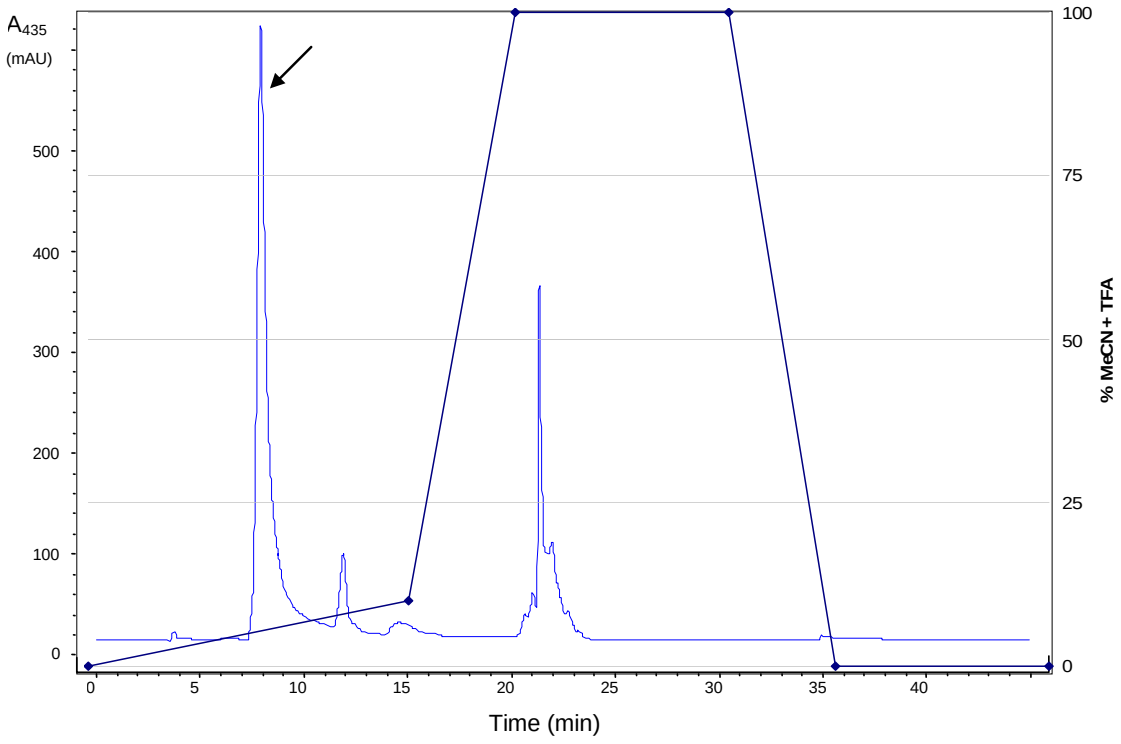


Figure 2-4: HPLC chromatogram of the separation of ferri-coelichelin from concentrated culture supernatants of *S. coelicolor* M145 overlaid with the H_2O (0.1% TFA)/MeCN (0.1% TFA) gradient used.

The arrow indicates the peak corresponding to ferri-coelichelin.

2.2.3 : Formation of desferri-coelichelin

All glassware was rinsed with concentrated nitric acid followed by deionised water to remove any metal ion contaminants which may associate with the siderophore. To form desferri-coelichelin, Fe(III) was removed by using 8-hydroxyquinoline as a chelating agent (Figure 2-5)[52,75].

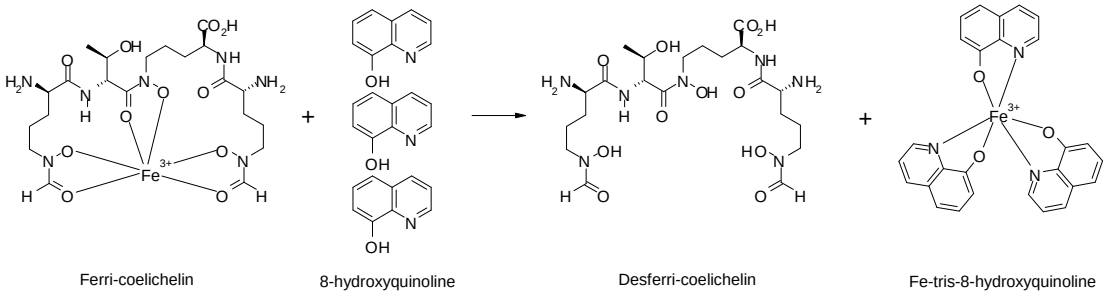


Figure 2-5: The scheme of the removal of Fe(III) from ferri-coelichelin by addition of 8-hydroxyquinoline.

A large excess of 8-hydroxyquinoline was dissolved in methanol and added to a dilute aqueous solution of ferri-coelichelin in water. The dark-green Fe(8-hydroxyquinoline)$_3$ complex was extracted using dichloromethane and the colourless aqueous phase was evaporated and lyophilized.

2.2.4 : RP-HPLC of desferri-coelichelin

As desferricoelichelin is unstable, this step was done without delay. The HPLC method of Lautru *et al.*[52] was used to purify desferri-coelichelin, monitoring the absorbance at 210 nm. Fractions that contained the m/z 566 (desferricoelichelin + H^+) and m/z 588 (desferricoelichelin + Na^+) were kept on ice after elution. The retention time of desferri-coelichelin was much shorter (7 minutes) than that reported in a previous purification (26 minutes)[76]. The combined fractions were snap-frozen in liquid N_2 and lyophilized.

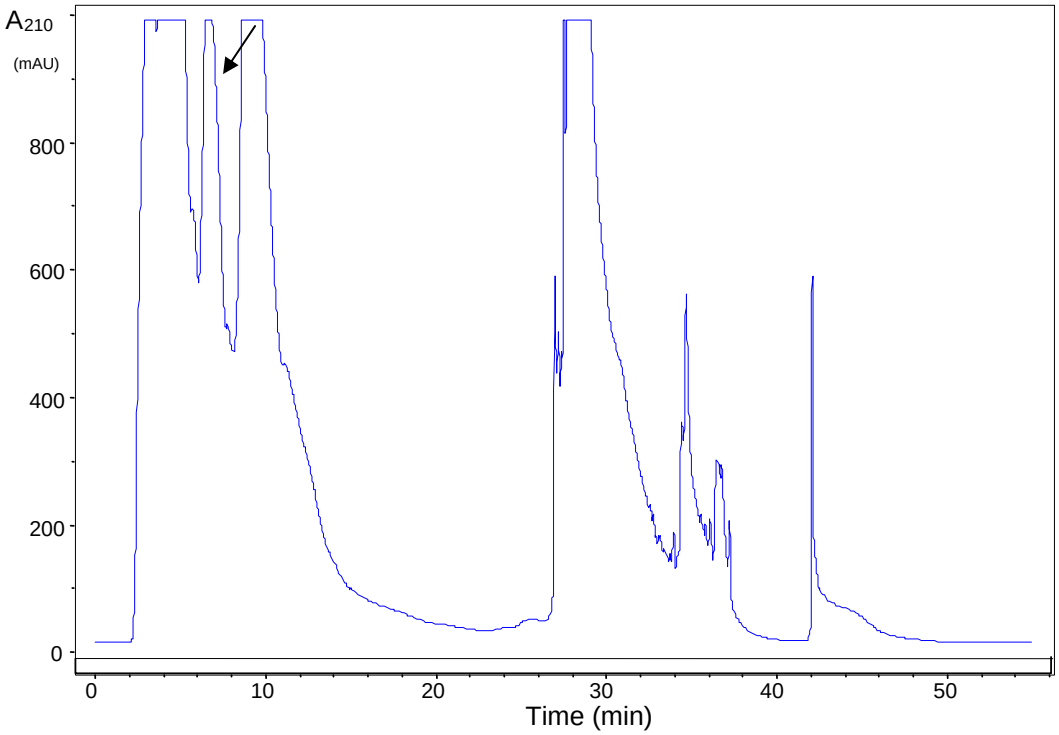


Figure 2-6: HPLC chromatogram of the purification of desferri-coelichelin, monitoring the absorbance at 210 nm.

The desferri-coelichelin peak is indicated by an arrow.

2.2.5 : Formation of gallium-coelichelin

One of the most conclusive ways to determine the organic purity of a compound is to measure a nuclear magnetic resonance (NMR) spectrum of it. However, the ferric complex of coelichelin contains paramagnetic high-spin Fe(III), which causes excessive broadening of the NMR spectra. To form a stable complex which could be analysed by NMR spectroscopy, the gallium complex was formed using the published protocol[52]. Gallium sulfate in dilute sulfuric acid was mixed with a desferricoelichelin solution and allowed to react for 30 minutes. The solution was neutralized with NaOH and lyophilized in 15 mL high-sided centrifuge tubes (because the gallium-coelichelin powder produced was very labile and was sucked out of round-bottom flasks by the low pressure conditions of the freeze-drier).

2.2.6 : RP-HPLC purification of gallium-coelichelin

The gallium-coelichelin complex was dissolved in 10 mM ammonium carbonate buffer and purification was carried out using the same conditions that was used to separate desferri-coelichelin (see above), but no fraction corresponding to gallium-coelichelin was obtained. Instead, a semi-preparative Zorbax C-18 column was used. The fractions containing gallium-coelichelin (Figure 2-7) was detected by positive ion ESI-TOF-MS in positive ion mode by the presence of m/z 632 (gallium-coelichelin + H^+). The fractions were combined and lyophilized.

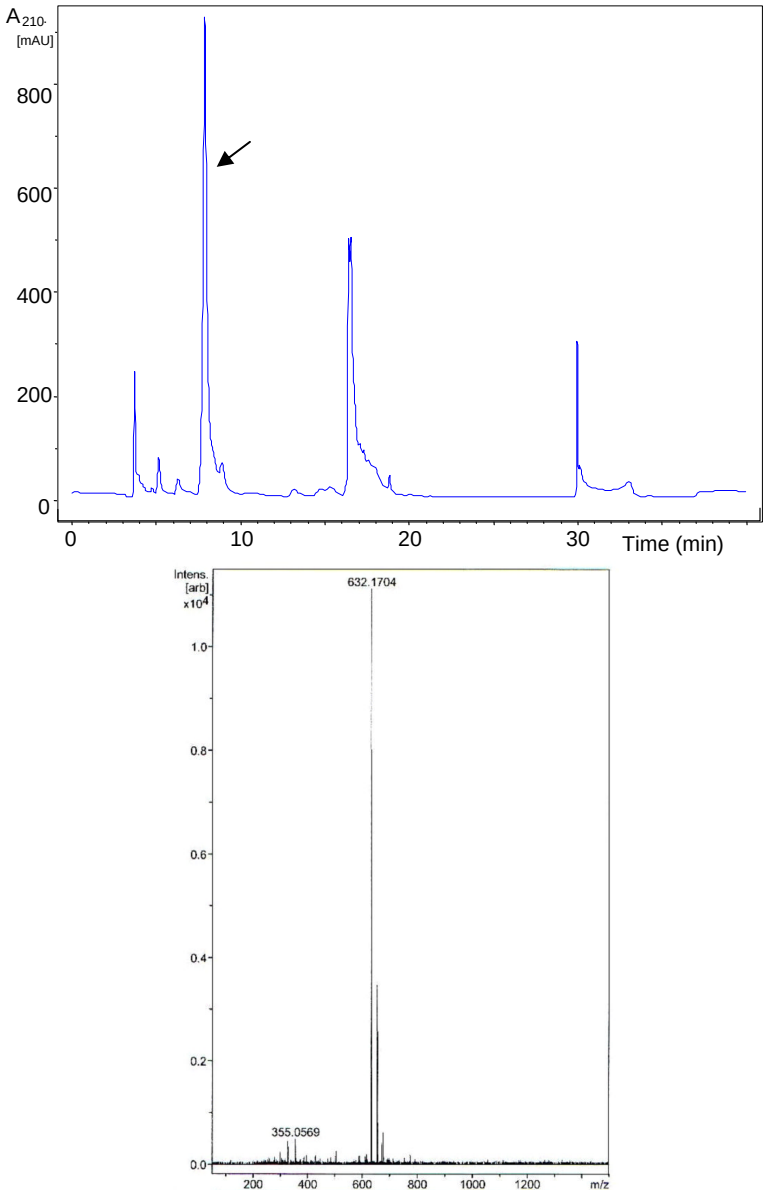


Figure 2-7: HPLC chromatogram of the separation of gallium-coelichelin, with absorbance monitored at 210 nm.

The gallium-coelichelin peak is indicated by an arrow. The presence of gallium-coelichelin was confirmed by ESI-TOF-MS (bottom) from the presence of the peaks m/z 632 (gallium-coelichelin + H^+) and 654 (gallium-coelichelin + Na^+).

2.2.7 : ^{1}H NMR analysis of gallium-coelichelin

Ga-coelichelin was dissolved in DMSO-d_6, a polar solvent which does not exchange deuterons for protons, and a ^{1}H NMR spectrum was recorded at 700 MHz. The most apparent features of the spectrum were large peaks corresponding to the protons from ammonium carbonate (7.25 ppm; from the HPLC liquid phase) and residual water (3.25 ppm) respectively, plus three sharp peaks corresponding to DMSO-d_6 (2.34, 2.44, 2.53 ppm). Most of the other peaks correspond to those reported previously for gallium-coelichelin[52] and are listed in Table 2-1. COSY and HSQC spectra of gallium-coelichelin were used to assist in assigning the ^{1}H NMR spectrum.

2.2.8 : Estimation of concentration of gallium-coelichelin

Prior to addition of $FeCl_3$ to reform the Fe-coelichelin complex, the concentration of gallium-coelichelin needed to be estimated. To do this, a known concentration of dibromomethane was added to the NMR tube containing the purified gallium-coelichelin and the NMR spectrum was measured. The integral of the new peak for the two protons of CH_2Br_2 (5.35 ppm) was compared with the integral for the proton attached to the α-carbon of the L-Thr residue of gallium-coelichelin (4.59 ppm).

2.2.9 : Addition of iron to gallium-coelichelin

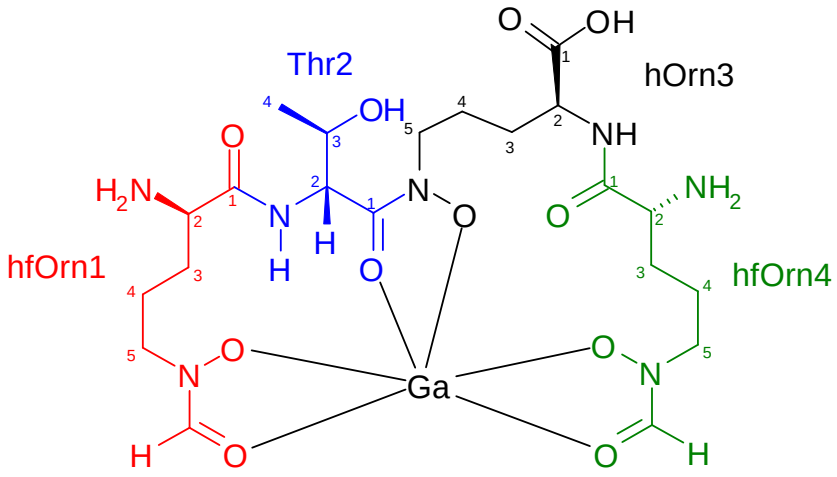
As coelichelin is a siderophore, it is a high-affinity chelator of ferric iron with greater affinity for this ion than for other ions such as Ga(III). To form the ferric complex, an excess of $FeCl_3$ (1 mM) was added and allowed to react for 12 hours to displace the Ga(III) ions with Fe(III) ions, using the estimated concentration of Ga-coelichelin from the NMR analysis to determine the amount of $FeCl_3$ solution to add. The conversion was monitored by positive-ion ESI-TOF-MS analysis (Figure 2-8).

The disappearance of the species with m/z 632 (gallium-coelichelin + H^+) and the appearance of a species with m/z 619 (ferri-coelichelin + H^+) was observed.

Table 2-1: (^{1}H)NMR assignments of gallium-coelichelin in Lautru *et al.*[52] and this study. Analysis of the COSY and HSQC spectra was also used to aid assignment of the resonances. The letters m, dd, d and s denote multiplet, double doublet, doublet and singlet peaks respectively.

Assignment	δ_H (Lautru *et al.*[52])	δ_H (this study)
hfOrn1-C1	-	-
hfOrn1-C2	3.97 dd, 3.0 Hz, 3.0 Hz	3.91 dd, 3.0 Hz, 3.0 Hz
hfOrn1-C3	1.91 m; 1.50 m	1.86 m; 1.40 m
hfOrn1-C4	1.42 m; 1.50 m	1.41 m; 1.50 m
hfOrn1-C5	3.41 m; 3.51 m	3.45 m; 3.56 m
hfOrn1-NH_2	-	-
hfOrn1-C(O)H	8.17 s	8.20 s
Thr2-C1	-	-
Thr2-C2	4.59, 9.0 Hz, 9.0 Hz	4.59, 9.0 Hz, 9.0 Hz
Thr2-C3	3.82 m	3.84 m;
Thr2-C4	1.14 d, 5.5 Hz	1.17 d, 7 Hz
Thr2-NH	8.75 d, 9.0 Hz	8.79 d, 9.0 Hz
Thr2-OH	5.08	5.08
hOrn3-C1	-	-
hOrn3-C2	4.24 dd, 8.5 Hz, 11.0 Hz	4.22 dd, 11 Hz, 10 Hz
hOrn3-C3	1.76 m	1.76 m
hOrn3-C4	1.83 m; 1.83 m	1.85 m; 1.84 m
hOrn3-C5	3.45 m; 3.93 m	3.47 m; 3.90 m
hOrn3-NH	7.57 d, 8.5 Hz	7.70 d, 8.0 Hz

hfOrn4-C1	-	-
hfOrn4-C2	4.15 dd, 11.0 Hz, 1.9 Hz	4.05 dd, 11.4 Hz, 2.0Hz
hfOrn4-C3	1.64 m; 2.29 m	1.41 m, 2.17 m
hfOrn4-C4	1.96 m; 1.76 m	1.94 m; 1.76 m
hfOrn4-C5	3.76 m	3.73 m
hfOrn4-C(O)H	8.22 s	8.22 s
hfOrn4-NH_2	-	-



2.3 : LC-MS and ESI-MS-MS characterization of purified ferri-coelichelin

2.3.1 : LC-MS characterization of purified ferri-coelichelin

Using the conditions of Barona-Gomez *et al.*[53], the purified ferri-coelichelin was analysed by LC-MS to determine if it has a similar retention time to ferri-coelichelin in concentrated *S. coelicolor* M145 culture supernatants. The retention time for purified ferri-coelichelin (3.1 minutes) was very similar to the retention time of ferri-coelichelin from *S. coelicolor* M145 and W2 supernatants (3.2 minutes) (Figure 2-9).

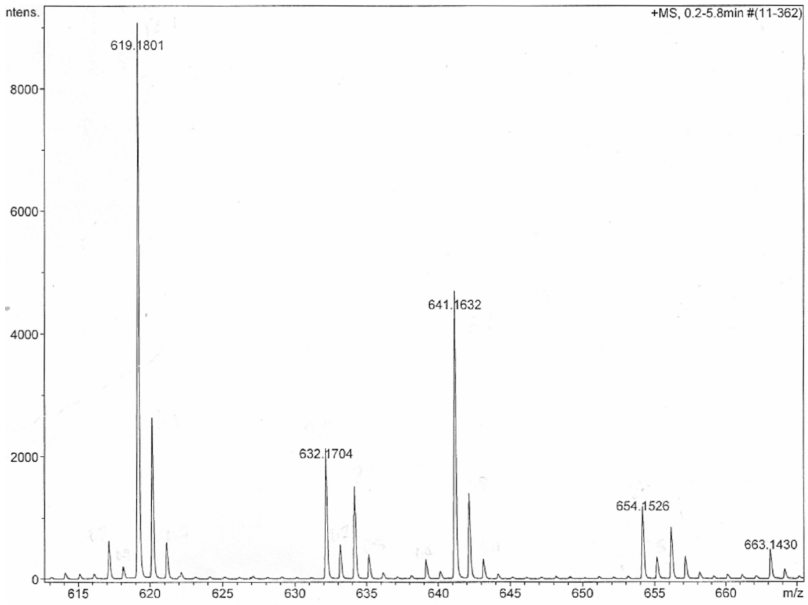


Figure 2-8: ESI-MS spectrum for a partially-ferrated sample of gallium-coelichelin.

The peaks with m/z 632 (Ga-coelichelin + H^+) and 654 (Ga-coelichelin + Na^+) appear with the larger ferri-coelichelin peaks with m/z 619 (Fe-coelichelin + H^+) and 641 (Fe-coelichelin + Na^+). The ferri-coelichelin conversion was complete (no m/z 632 peak) on addition of an excess of $FeCl_3$.

2.3.2 : UV-Vis spectrum of ferri-coelichelin

The UV spectrum corresponding to the m/z 619 peak in the pure sample of ferri-coelichelin was analysed in the HPLC-MS diode array detector from 250 nm to 800nm. The compound had a large absorbance at the violet - ultraviolet end of the spectrum at 425 nm, close to the 435 nm maximal absorbance distinctive of other ferric-*tris*-hydroxamates (Figure 2-10).

2.3.3 : Positive ion mode ESI-MS-MS of characterization of purified ferri-coelichelin

When pure ferri-coelichelin was analysed by positive ion mode ESI-MS, several peaks were observed. As well as the peaks with m/z 619 (ferri-coelichelin + H^+), 641 (ferri-coelichelin + Na) and 663 (ferri-coelichelin + $2Na^+$), there were peaks at m/z 557, 575, 597 and 601(Figure 2-11).

To investigate whether these were contaminants in the sample or were fragment ions of ferri-coelichelin, ESI-MS-MS was performed on the m/z 619 peak (Figure 2-12), which yielded peaks with m/z 601, 583, 575, 557 and 540, which all of them can be attributed to fragment ions of ferri-coelichelin (Table 2-2).

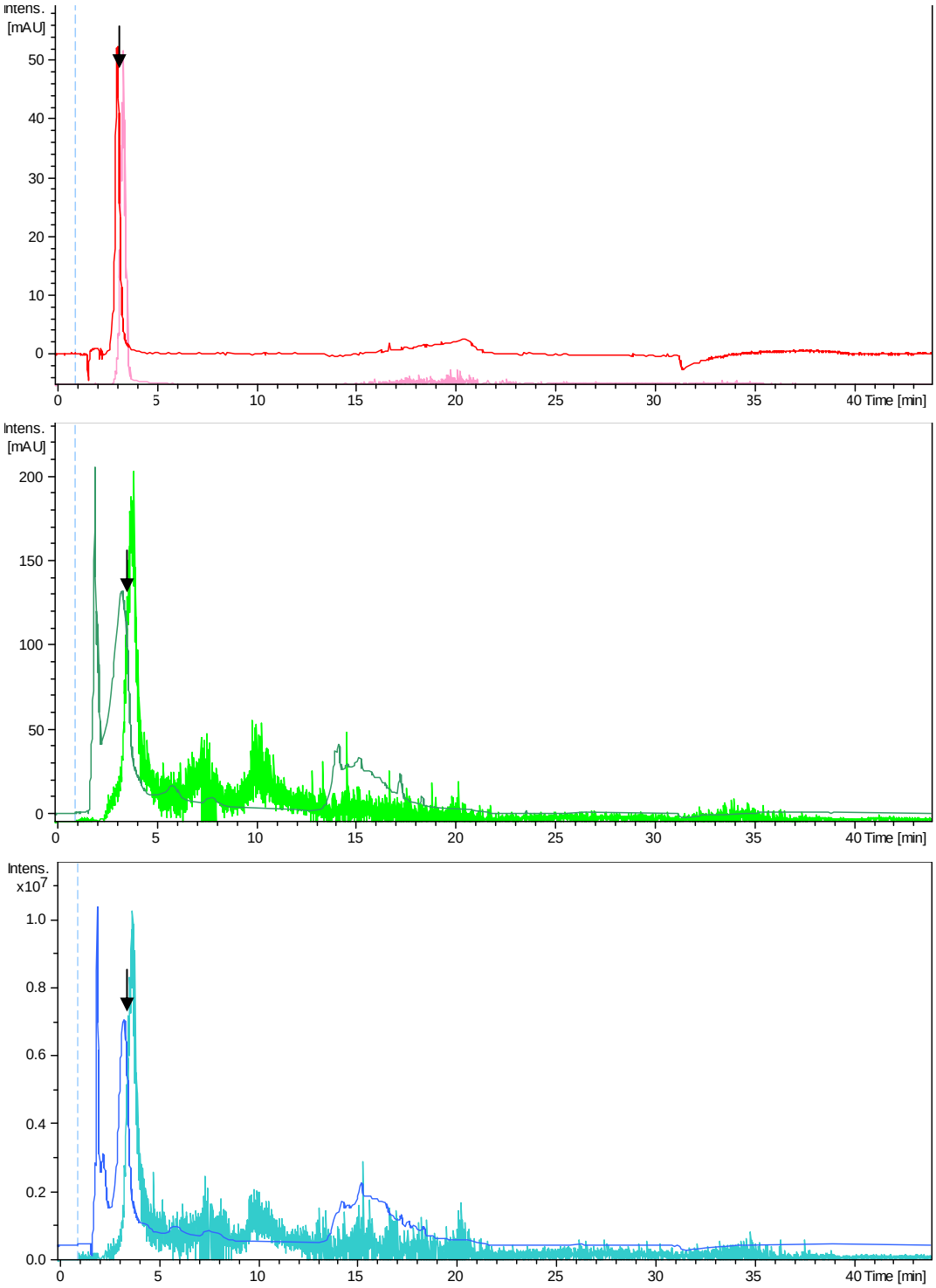


Figure 2-9: LC-MS analysis of purified ferri-coelichelin and ferri-coelichelin in *S. coelicolor* M145 and W2 culture supernatants.

UV chromatograms at 435 nm from HPLC-MS analysis of purified ferri-coelichelin (top, red) plus the $FeCl_3$-supplemented supernatants of cultures of *S. coelicolor* M145 (middle, green) and W2 (bottom, blue). The chromatograms are overlaid with their corresponding extracted ion chromatograms (light colours). The arrows indicate the peak in the UV chromatograms that correspond to ferri-coelichelin.

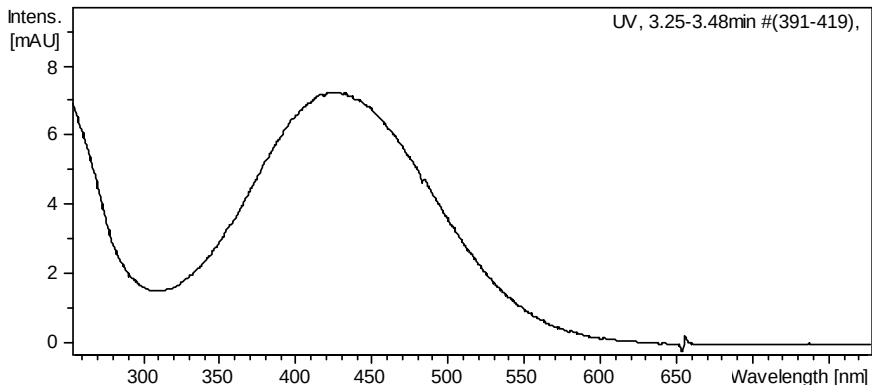


Figure 2-10: UV-Vis spectrum of the ferri-coelichelin peak from 3.25 - 3.48 min from the LC-MS analysis of purified ferri-coelichelin in Figure 2-9.

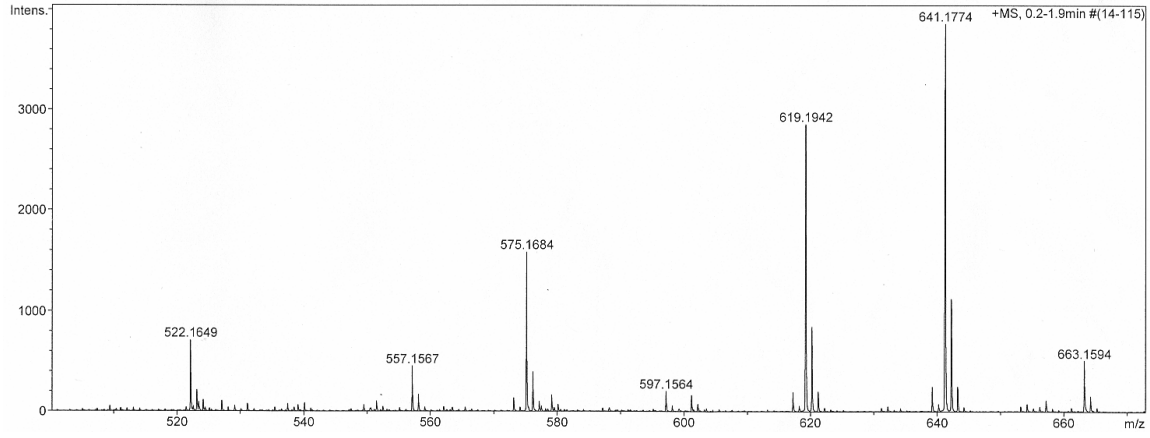


Figure 2-11: Positive ion mode ESI-MS of a dilute sample of purified ferri-coelichelin in Bruker micrOTOF.

The presence of 619 m/z, 641 m/z and 663 m/z were indicative of unfragmented ferri-coelichelin, whilst the other peaks (522 m/z, 557 m/z, 575 m/z, 597 m/z) were indicative of fragmented ferri-coelichelin, which was confirmed by ESI-MS-MS on the same sample.

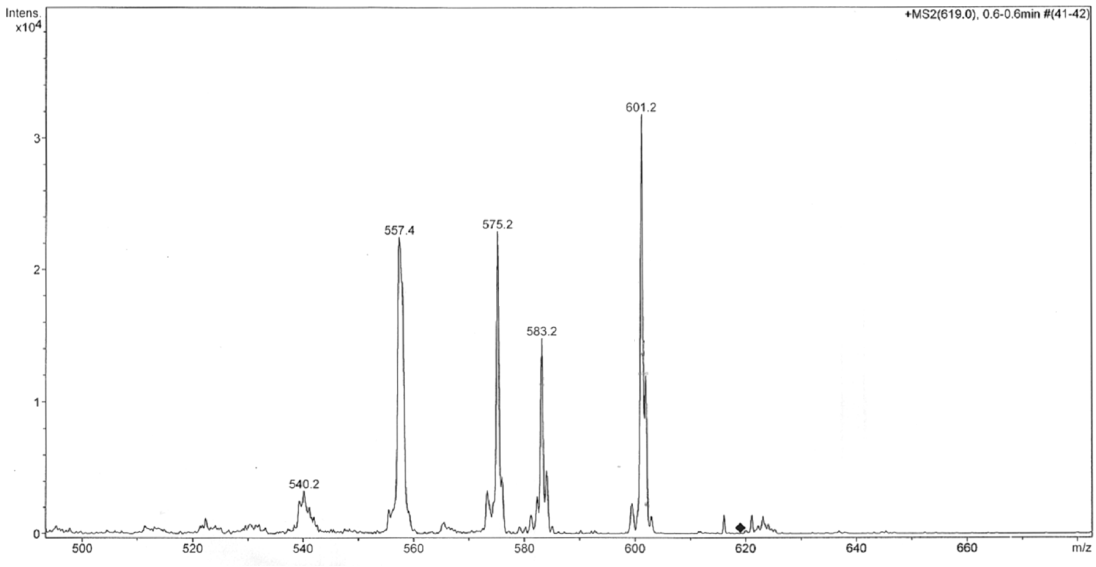


Figure 2-12: ESI-MS-MS spectrum of purified ferri-coelichelin (m/z = 619; m+H^+).
The fragmentation pattern of m/z 619 was attributed to the loss of H_2O in hydroxamate groups and the loss of CO_2 from a carboxyl group (Table 2-2).

Table 2-2: Possible fragmentation of ferri-coelichelin

m/z	Possible ion
619	Fe-cclin + H^+
601	Fe-cclin + H^+- H_2O
583	Fe-cclin + H^+- $2H_2O$
575	Fe-cclin + H^+- CO_2
557	Fe-cclin + H^+- CO_2 - H_2O
540	Fe-cclin + H^+- CO_2 - $2H_2O$

Chapter 3: Overproduction, isolation and characterization of recombinant proteins DesE, CchF and CdtB

3.1 : Introduction

To elucidate unequivocally the routes of ferri-siderophore transport in *Streptomyces coelicolor*, one must study the lipoprotein receptors, the high-affinity siderophore binding proteins that govern ferri-siderophore entry. Given the low concentration of ferri-siderophores in solution, their binding to the lipoprotein receptors are likely to be the rate-limiting step in their absorption. The genes encoding the lipoprotein receptors also have to be expressed under iron limitation. The best indicator of this is if they are flanked by or adjacent to iron boxes, palindromic sequences to which the *S. coelicolor* Fe(II)-dependent repressors DmdR bind.

Searching the *S. coelicolor* genome for sequences similar to the iron box upstream of the gene *desE* (TTAGGTTAGGCTCACCTAA)[20], the only lipoprotein receptor genes that were near this or similar sequence of bases were *cchF*, *cdtB*, *desE* and SCO2272. In a study of the membrane-associated proteome of *S. coelicolor*, only the proteins CchF, CdtB and DesE proteins were identified[62], and have been postulated to play a role in siderophore uptake. Although one cannot rule out the role of other *S. coelicolor* genes (SCO2272, SCO0996 and SCO7218) in ferric siderophore transport, DesE, CchF and CdtB were the best candidates for the binding and subsequent transport of the cognate siderophores of *S. coelicolor.*

The general approach was to overproduce the proteins encoded by the *S. coelicolor* A3(2) genes *desE*, *cchF* and *cdtB* in *Escherichia coli* for binding studies, with an N-terminal His-tag to assist with purification with immobilized metal-affinity chromatography (IMAC), which is a common strategy for purification of many bacterial proteins[77].

3.2 : Features of the plasmid used for recombinant protein production

For recombinant protein overproduction in *E. coli*, the plasmid pET151-D/TOPO from Invitrogen[78] was used to overexpress of the *desE, cchF* and *cdtB* genes. This plasmid has many attractive features which made it useful for protein overexpression.

The gene of interest is amplified by polymerase chain reaction (PCR) using primers complementary to the 5' and 3' end of the gene of interest, but with the bases CACC appended to the 5' end. This series of bases in the insertion DNA is complementary to the GTGG sticky end of the plasmid to give directional insertion into the plasmid, via the enzyme Topoisomerase I (from *Vaccinia* virus). This system avoids the multiple DNA ligation reactions usually involved in gene insertion into plasmids. This reaction takes only 15 minutes and the product is used immediately to transform competent strains of *E. coli*.

Once used to transform *E. coli*, the plasmid product is able to undergo low-copy replication by the presence of the pBR3222 *ori* (origin of replication) and *rop* genes (Figure 3-1). The *E. coli* cells that had been successfully transformed with the plasmid can be selected with the antibiotic ampicillin as the plasmid has the *bla* gene (encodes β-lactamase) which confers resistance to this antibiotic.

When the plasmid is used to transform strains of *E. coli* containing T7 RNA polymerase, such as the overproducing strain BL21star (DE3), transcription of the gene is able to take place. This is regulated by the *lac* operon, in which the *lacI* gene encodes a repressor which binds onto the *lac* operator region, but the addition of a lactose analogue such as isopropyl β-D-1-thiogalactopyranoside (IPTG) reduces the affinity of the LacI protein for the *lac* operator, enabling T7 polymerase to bind to the T7 promoter site. RNA transcripts from the T7 promoter site to the T7 termination site (T7 term) are produced. The T7 promoter is a strong promoter, enabling the production of a large number of RNA transcripts for high-level protein overproduction.

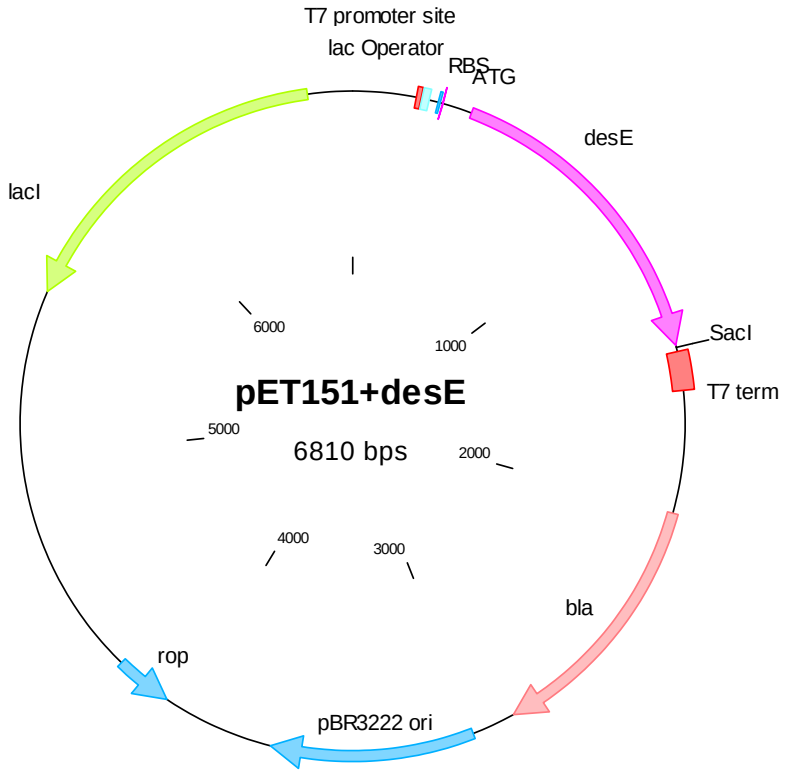


Figure 3-1: Diagram of pET151 plasmid with the full-length *desE* gene.

The *desE* insert encodes the protein desired to be overproduced when this plasmid is used to transform *E. coli.* The regions of the plasmid are explained in the text.

E. coli ribosomes are able to bind to the ribosome binding site (RBS) of the RNA transcripts and translation is initiated at the start codon ATG to the stop codon of the inserted gene. Another useful feature of the plasmid is that it encodes the protein to have an N-terminal hexahistidine tag for purification (Figure 3-2) by immobilized metal affinity chromatography (IMAC) by Ni-NTA column. The consecutive histidine residues form coordinate bonds with Ni^{2+} which immobilizes the protein to the column. The protein is eluted by addition of high concentrations of imidazole, a histidine side-chain analogue, which also forms stable coordinate bonds with Ni^{2+}. After elution, this tag can be cleaved off the recombinant protein by tobacco etch virus (TEV) protease in a further purification step, leaving an N-terminal tag of six residues plus the protein sequence desired (Figure 3-2).

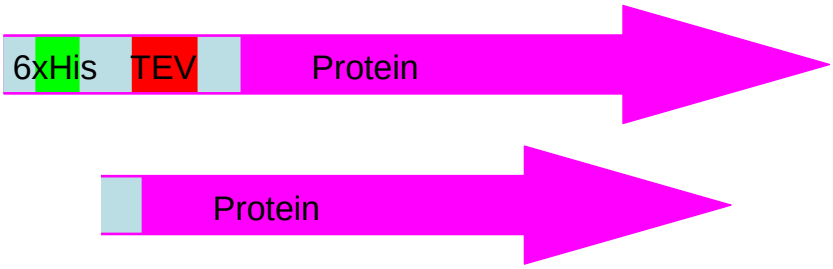


Figure 3-2: Expected motifs of proteins overproduced using the pET151/D-TOPO plasmid.

(Top) The overproduced protein will have N-terminal hexahistidine tag (6xHis) and TEV protease cleavage site (TEV site) and a spacer (light blue) between the site and the native protein sequence (pink). (Bottom) After digestion by TEV protease, most of the N-terminal tag of the protein is cleaved off, leaving the required protein sequence plus a 6-residue N-terminal tag (GIDPFT).

3.3 : Overproduction and purification of full-length DesE, CchF and CdtB

3.3.1 : PCR primer design of *desE*, *cchF* and *cdtB* genes

Initially, it was decided the proteins were to be overproduced as full-length proteins, using the complete gene sequence. They have an N-terminal signal sequence encoded which is recognised by translocases which transport them to the outer membrane of *S. coelicolor*. The protein DesE has been shown to be a substrate by the *S. coelicolor* Tat translocase system[63], and due to the presence of N-terminal twin-arginine motifs in CchF and CdtB they may also be transported by this system[63].

Sequences for the *S. coelicolor* genes *desE* (SCO2780), *cchF* (SCO0494) and *cdtB* (SCO7399) were obtained from the annotated genome sequence[51] accessed via Entrez NCBI (http://nih.ncbi.gov/Entrez). All primers were designed using the software Clone Manager Suite[79]. Forward primers were designed such that they matched the first 15 – 18 base pairs in the database sequence plus the bases CACC at the 5' end. Reverse primers were designed such that they matched the last 15 – 18 base pairs of the complementary strand of the gene sequence.

3.3.2 : Analysis and purification of PCR product

PCR products were produced as described in the materials and methods (section 8.3.3 :). A strong band at around 1 kbps (kilo base pairs) was indicative of a successful PCR reaction (Figure 3-3). The expected size of each amplimer is given in Table 3-1. The concentration of each amplimer was estimated by comparing the intensity of the bands of known concentration in the DNA size marker. A larger amount of amplimer solution PCR was run on the gel to separate the desired PCR product from unused reactants and template DNA. The 1 kbps band was excised from the gel and subsequently purified before ligation into the expression plasmid.

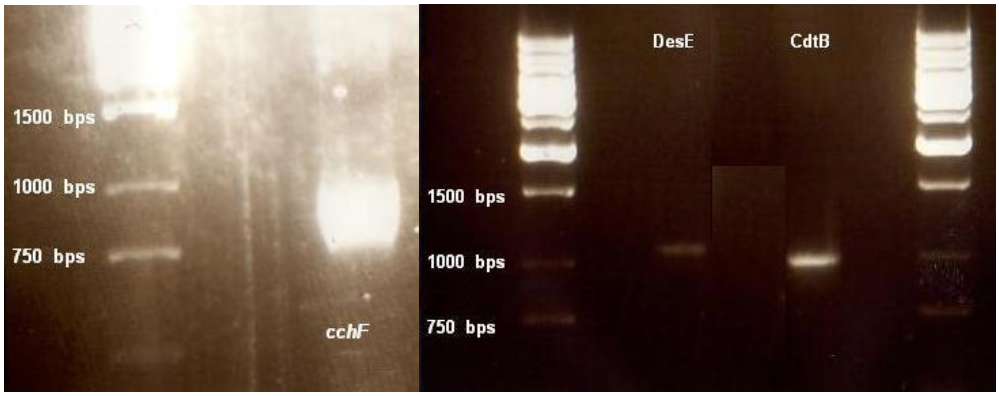


Figure 3-3: Agarose gel electrophoresis of full-length amplimers *cchF*, *desE* and *cdtB*

(Left) *cchF* product amplified by polymerase chain reaction (PCR) from *S. coelicolor* cosmid SCF94. (Right) The genes *desE* and *cdtB* amplified from *S. coelicolor M145* genomic DNA.

3.3.3 : Insertion of amplimers into the plasmid pET151/D-TOPO plasmid (TOPO reaction) and subsequent transformation of *E. coli*

The protocol described in the manual[78] was used to insert the DNA into the pET151/D-TOPO plasmids. The product was used to transform chemically competent *E. coli* TOP10 cells or *E. coli* MC1061 by electroporation.

3.3.4 : Analysis of the insert of *E. coli* clones by PCR screening

Eight to ten *E. coli* transformants were grown overnight for isolation and analysis of their plasmids, which were extracted using a standard procedure[80] and used as the

template DNA for PCR reactions using the PCR primers and conditions as in the original gene amplification reactions (but using a cheaper polymerase). The PCR products are analysed by agarose gel electrophoresis for a band at approximately 1 kbps (see Figure 3-13 for an example with truncated amplimers). This method, PCR screening, was an effective method for determining whether the gene has been inserted into the plasmid.

Table 3-1: Expected size of amplified genes (full-length)

Gene	Expected size (DNA base pairs)
desE	1050
cchF	1053
cdtB	984

3.3.5 : Analysis of overproduction of proteins

Plasmids containing the correct inserts were used to transform *E. coli* BL21star (DE3), an overproducing strain of *E. coli* containing T7 RNA polymerase. This strain is able to transcribe the insert region of the plasmid, hence producing the protein encoded within. Single colonies from overnight LB agar plates were grown overnight in liquid culture. This starter culture was used to inoculate a larger liquid culture (500 mL to 1 L), which was incubated at 37 °C until the cell density had reached OD_{600} = 0.6 - 0.8. IPTG was added later to induce transcription of the T7 polymerase inside the bacteria at a temperature of 15 °C for 15 – 20 hours. The total protein and soluble protein (cell-free extract) fractions of the cell paste were analysed by SDS-PAGE using Coomassie staining to determine if overproduction of the desired proteins had taken place, and if these proteins were soluble.

All three proteins, His_6-DesE, His_6-CchF and His_6-CdtB were found to be both overproduced and soluble in *E. coli* BL21star indicated by bands at around 40 – 45 kDa, but in the case of CchF, two proteins of a similar molecular weight were overproduced (Figure 3-4). The protein DesE was visible as a single band at 40 kDa, whilst the band corresponding to CdtB ran at almost 50 kDa, which was much larger than the predicted molecular weight of the protein.

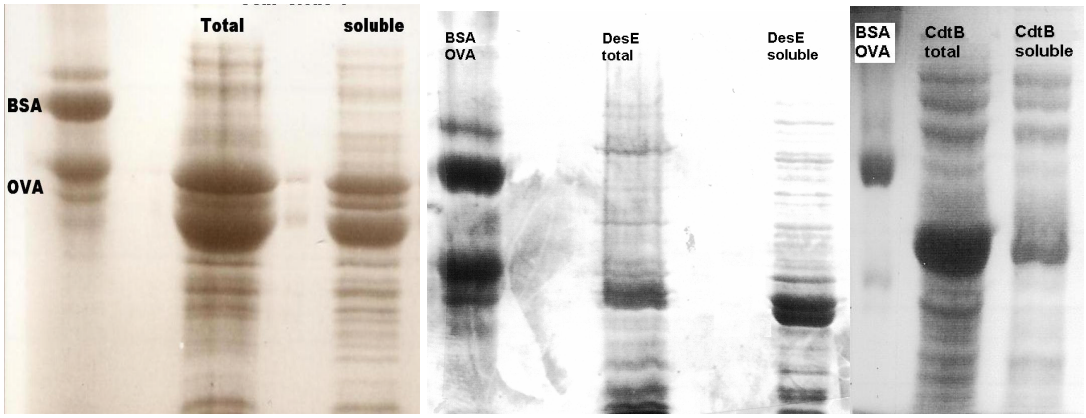


Figure 3-4: Overproduction of the full-length proteins His_6-CchF, His_6-DesE and His_6-CdtB in transformed clones of *E. coli* BL21star.

Photographs of 8 % SDS-PAGE gels of the total and soluble (cell-free extract) fractions of His_6-CchF (left), His_6-DesE (middle) and His_6-CdtB (right) overproducing clones of *E. coli* BL21star transformed with the plasmid pET151/D-TOPO, against BSA and OVA proteins, which were used as molecular weight standards.

Table 3-2: Expected protein sizes of His_6-DesE, His_6-CchF and His_6-CdtB

Protein	Size (Da)
His_6-DesE	40425.0
His_6-CchF	41157.8
His_6-CdtB	37877.1

3.3.6 : Large-scale purification of His_6-DesE, His_6-CchF and His_6-CdtB by Ni-NTA IMAC

The integrity of the plasmid clones resulting in protein overproduction was checked by DNA sequencing using T7 and T7 reverse primers. Clones without mutations or silent mutations were selected for large-scale protein production. The pET151/D-TOPO plasmids containing the full-length genes *desE*, *cchF* and *cdtB* will be denoted by the names given in Table 3-3. These plasmids were used to transform competent *E. coli* BL21star prior to protein overproduction.

About 600 mL of *E. coli* BL21star was grown and protein overproduction induced by addition of 0.5 mM IPTG and incubated at 15 °C overnight. The cell-free extract

from the cell paste was obtained from lysis by French Press, and proteases were inhibited by phenylmethylsulfonylfluoride (PMSF).

Table 3-3: Names of pET151 plasmids containing the full-length genes *desE*, *cchF* and *cdtB*.

Plasmid name	Gene inserted into pET151/D-TOPO plasmid
pPP001	*desE*
pPP002	*cchF*
pPP003	*cdtB*

Ni-NTA IMAC (immobilized metal affinity chromatography) is a very common method for the purification of proteins which have a polyhistidine tag engineered onto their N-terminus or C-terminus. At these positions the protein is flexible, where side chains of contiguous six histidine residues to form a stable complex with Ni(II) immobilized on a nitriloacetic acid (NTA) modified sepharose column. This binding is greatest at alkaline pH.

To remove proteins that bind non-specifically to the column, it is washed with a buffer containing a low concentration of imidazole (around 10 mM – 20 mM). This compound is chemically similar to the histidine side chains and also forms complexes with Ni(II). To elute the protein from the column, a buffer containing a high concentration of imidazole (300 mM) is used, which competitively displaces the His-tagged protein from the material. The eluted protein is filtered and concentrated to remove the high concentrations of imidazole using a centrifugal filter, and stored in a glycerol containing buffer at -80 °C.

Purification of His_6-CchF from the cell-free extract from *E. coli* BL21star overproducing cells (*E. coli* BL21star transformed with the plasmid pPP002) by IMAC was attempted unsuccessfully. The large band corresponding to the protein product as seen in the cell-free extract was not present in the IMAC elution, indicating the protein was not binding to the Ni-NTA column and washed off (Figure 3-5). The protein bands on the SDS-PAGE were analysed by peptide mass fingerprinting, which confirmed they are derived from the CchF-encoding sequence.

A likely explanation for this is that CchF contains a twin-arginine translocase[64] (TAT) signal in its N-terminus, which causes CchF to be transported to the outside surface of the membrane in *S. coelicolor.* This signal sequence of full-length His_6-CchF could be recognised by the *E. coli* TAT system, hence it is transported to the periplasm of the bacteria (Figure 3-6). During the process of translocation, the signal peptide was cleaved by the system, including the His tag, producing a lower molecular weight protein which cannot be purified by IMAC.

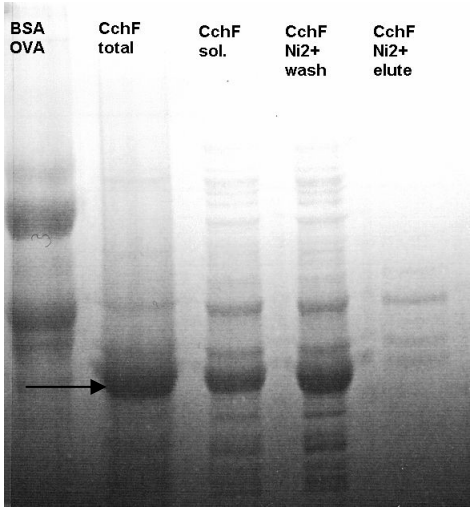


Figure 3-5: Attempted purification of the protein full-length CchF by Ni-NTA IMAC.

The main protein band (approximately 35 kDa; indicated by the arrow) is clearly visible in the total and soluble fractions of *E. coli* BL21star, and also appears in the cell-free extract after filtration in the Ni-NTA column (CchF Ni2+ wash), but elution of proteins bound to the column (with a buffer containing 300 mM imidazole), this band was very faint (CchF Ni2+ elute), implying this protein did not bind to the column hence not have a His-tag. The other bands in this fraction may correspond to non-specific proteins.

Purification of His_6-DesE and His_6-CdtB from *E. coli* BL21star transformed with the pPP001 and pPP003 plasmids respectively, was less problematic, although the cell pellet of the strain of *E. coli* overproducing His_6-CdtB had to be frozen overnight at -20 °C. Both proteins were overproduced, soluble and purified by Ni^{2+} IMAC, with yields of around 4 mg per litre of culture (determined by Bradford assay [81]). Analysis of purified His_6-DesE and His_6-CdtB by SDS-PAGE (Figure 3-7) showed large bands at 38 kDa and 42 kDa respectively, plus a smaller band at roughly double the molecular weight.

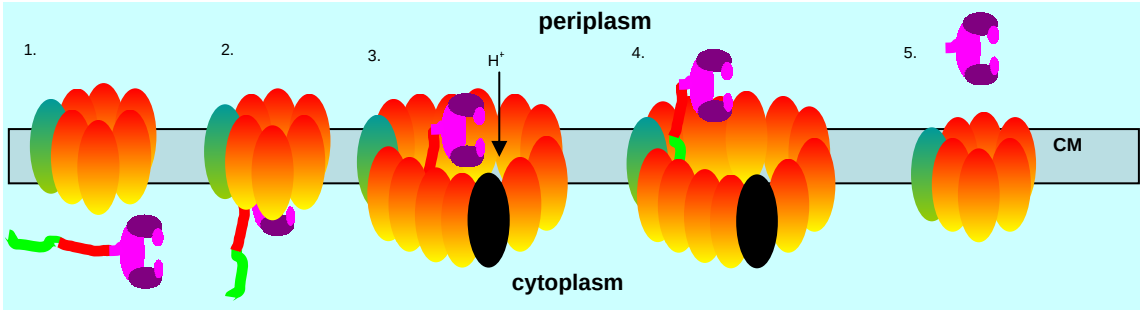


Figure 3-6: Proposed explanation for the cleavage of the His-tag in full-length CchF in overproducing cells of *E. coli*.

1. The protein (purple) is recognised as a Tat substrate by *E. coli* from its N-terminal signal peptide (red) with His-tag (green) and soon after translation is chaperoned to the cytoplasmic membrane (CM). 2. The proteins in the Tat complex TatBC (blue ovals) and TatA (red ovals) start the translocation process. 3. More TatA proteins are recruited in the translocation of CchF, and the translocation is driven by the proton motive force difference between the cytoplasm and the periplasm. 4. The signal peptide of CchF is cleaved during or after translocation. 5. CchF resides in the periplasm of the *E. coli* cells, *sans* His-tag and signal peptide.

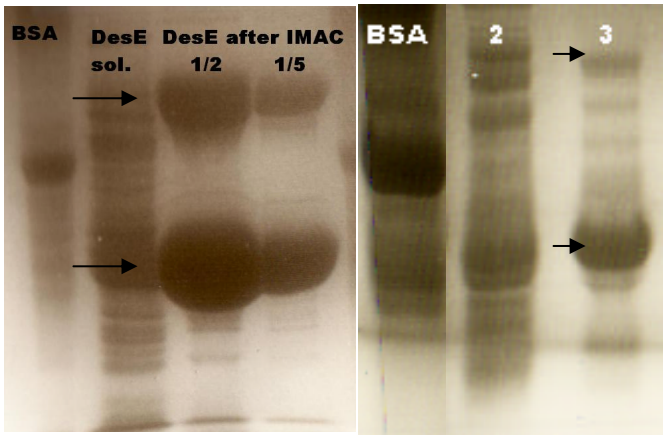


Figure 3-7: Purification of full-length His_6-DesE and His_6-CdtB by Ni-NTA IMAC.

(Left) The soluble fraction revealed the presence of a single overproduced protein (His_6-DesE indicated by arrow) of approximately 42 kDa (DesE sol.) but after Ni-NTA IMAC purification two large bands were seen (DesE after IMAC). (Right) Overproduction of His_6-CdtB The soluble fraction revealed the presence of a single overproduced protein (His_6-CdtB; column 2) of approximately 42 kDa, but after IMAC purification two bands were seen (column 3; arrows).

3.3.7 : Gel filtration of full-length DesE and CdtB

In an attempt to separate the lower molecular weight protein from the higher molecular weight protein, gel filtration (size exclusion chromatography) was used. In the case of full-length His_6-DesE (Figure 3-8), two major peaks were observed in the UV chromatogram with retention volumes of 56.4 mL and 63.8 mL. The retention time of the smaller peak corresponded to a threefold difference in the molecular weight of the larger peak, at 123 kDa and 42 kDa respectively. Fractions were collected at the peaks and analysed by SDS-PAGE. Both peaks corresponded to the lower molecular weight band, but the higher molecular weight band reappeared after concentration of the protein. Gel filtration of full-length His_6-CdtB (Figure 3-9) also had two large peaks as with His_6-DesE, but in this case all fractions contained both the higher and lower molecular weight protein by analysis by SDS-PAGE.

The most likely explanation for the gel filtration results of full-length His_6-DesE and His_6-CdtB is that these proteins oligomerize at high concentrations encountered during the protein purification process, with the dimer appearing even in the denaturing conditions of SDS-PAGE.

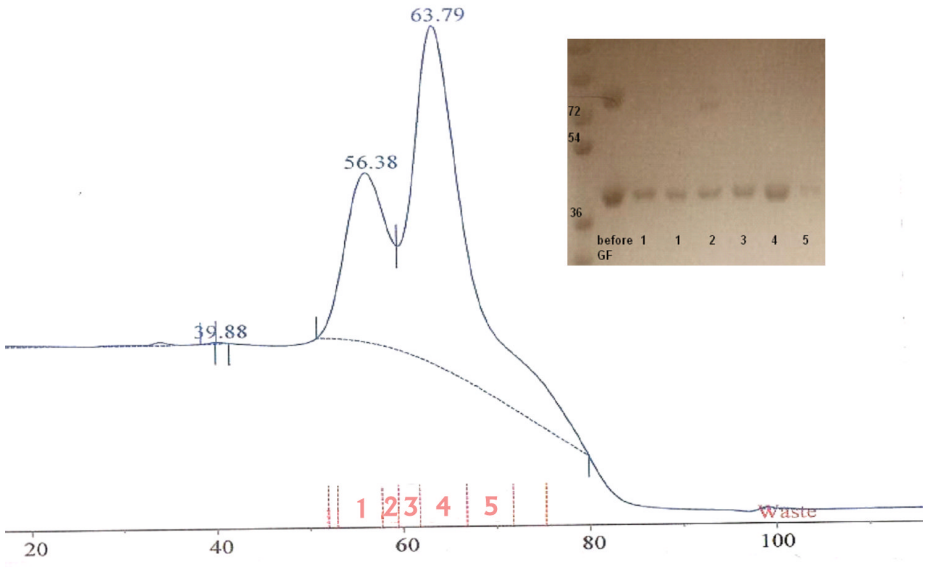


Figure 3-8: Gel filtration chromatograph of full-length His_6-DesE.

X-axis: UV absorbance at 280nm. Y-axis: Elution volume in mL. Two large peaks in the chromatograph (56.4 mL and 63.8 mL) were observed and fractions (1,2 and 4,5 respectively) were analysed by SDS-PAGE (inset) against His_6-DesE before gel filtration (before GF). Both peaks corresponding to different oligomers corresponded to the same band on SDS-PAGE.

To confirm the identity of His_6-DesE and His_6-CdtB, peptide mass fingerprinting was performed on upper and lower bands that appeared on the SDS-PAGE of purified His_6-DesE and His_6-CdtB, which confirmed they were the same protein.

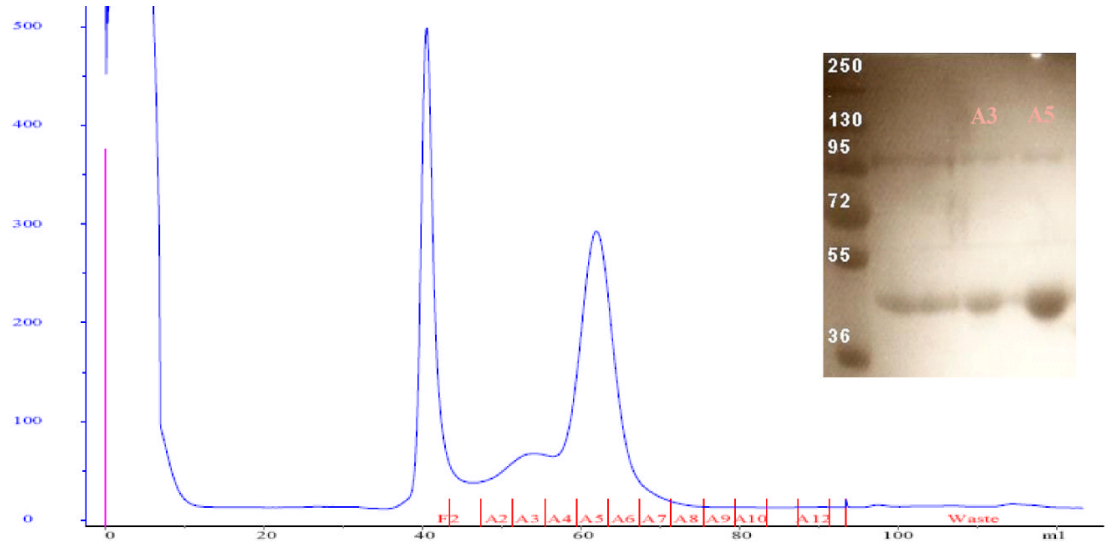


Figure 3-9: Gel filtration chromatograph of full-length His_6-CdtB.

X-axis: UV absorbance at 280nm. Y-axis: Elution volume in ml. Two large peaks in the chromatograph (53 mL and 63 mL) were observed and fractions (A3 and A5 respectively) were analysed by SDS-PAGE (inset). The A5 fraction had a higher proportion of lower molecular weight protein than the A3 fraction.

To obtain useful data from siderophores binding experiments, the proteins need to be a single oligomeric state. Dimer formation creates uncertainties about the location of the dimer interface and whether one or both siderophore binding sites are partially or fully observed. Thus we sought to overcome these problems by expressing truncated forms of DesE, CchF and CdtB lacking their hydrophobic N-termini, which we expected to be the cause of the oligomerization observed.

3.4 : Overproduction and purification of 5'-truncated DesE, CchF and CdtB

3.4.1 : Introduction

In previous studies of other Gram-positive siderophore lipoprotein receptors[34,38], recombinant proteins in which the N-terminus of the native protein sequences were truncated to the N-terminal cysteine, with this residue was mutated to a methionine. This removed the hydrophobic signal peptide which was remarked to be the cause of purification problems. These mutations are presumed to have little effect on binding site of the proteins as the signal peptide is not part of the mature protein in *S. coelicolor*, and the speculative role of the cysteine residue is as the anchor of an attached lipid which tethers the protein to the outer membrane (hence not involved in substrate binding). The proposed course of action was to truncate the DesE, CchF and CdtB proteins and the cysteine residue mutated to methionine to solve the purification problem (Figure 3-10).

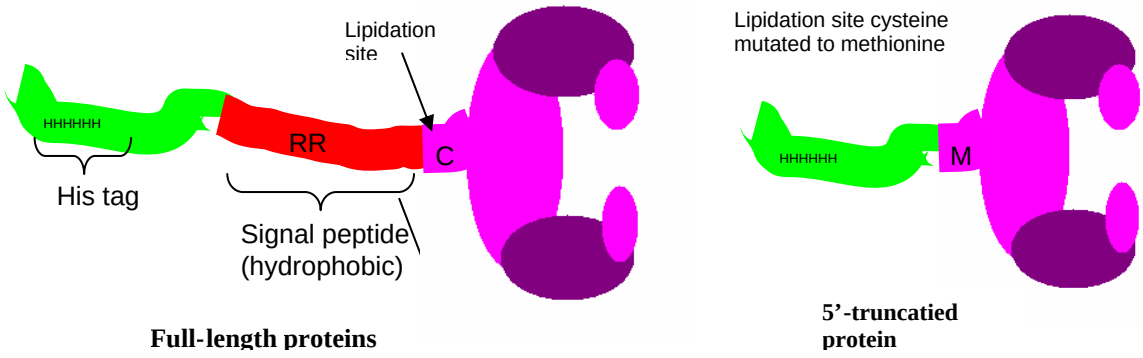


Figure 3-10: Proposed N-terminal modification of the full length proteins to produce the 5'-truncation proteins.

The native N-terminus containing the hydrophobic signal peptide (red) and the putative lipidation site cysteine of the full-length proteins are assumed to play no role in the globular domain involved in binding to ferri-siderophores (pink and purple). The proposed modification to produce the 5'-truncation proteins is to replace the signal sequence with the N-terminal tag as encoded in the pET151/D-TOPO with the His-tag (green) and to mutate the cysteine to methionine.

3.4.2 : PCR primer design for the 5'-truncated *desE*, *cchF* and *cdtB*

The reverse primers used in the amplification of the 5'-truncated genes (*desE*, *cchF* and *cdtB* reverse) were the same as those used in the amplification of the full length genes. The forward primers used were designed such that their 5' ends corresponded to the first cysteine codon that was predicted to be the lipid attachment site. This codon was mutated to a methionine codon (ATG) and preceding these bases CACC for correct insertion into the pET151/D-TOPO plasmid, giving the primers a CACCATG overhang at the 5' end (Figure 3-11).

3.4.3 : Amplification of genes by PCR

This was performed as for the full length *desE*, *cchF* and *cdtB* except as template DNA the plasmids that were used to overproduce the full length proteins (i.e. the plasmids pPP001, pPP002 and pPP003 respectively). The expected sizes of the amplimers are given in Table 3-4. The concentrations of the amplimers were estimated by comparing the intensity of the band on an agarose gel in comparison with the intensity of an appropriate band from a DNA ladder. The PCR reactions were separated on an agarose gel to isolate the required amplimer from unused reactants and template DNA.

1. atgctcctcagaacaacgcgtacgaagccctggcgacgactcgcggcggccctgtccgcc
MetLeuLeuArgThrThrArgThrLysProTrpArgArgLeuAlaAlaAlaLeuSerAla
gccgcgctcggcgtcggactcctcgcggggtgcggatccgactcggacgacccggcg
AlaAlaLeuGlyValGlyLeuLeuAlaGlyCysGlySerAspSerAspAspProAla

2. atgctcctcagaacaacgcgtacgaagccctggcgacgactcgcggcggccctgtccgcc
MetLeuLeuArgThrThrArgThrLysProTrpArgArgLeuAlaAlaAlaLeuSerAla
gccgcgctcggcgtcggactcctcgccaccatgggatccgactcggacgacccggcg
AlaAlaLeuGlyValGlyLeuLeuAlaThrMetGlySerAspSerAspAspProAla
Forward Primer

1. First 117 bases of *cchF* gene with the 90 bases encoding the N-terminal sequence that was required to be removed from the 5'-truncated protein highlighted in light blue. The twin arginine residues distinctive of TAT (twin arginine translocase) signal peptides are highlighted in dark blue, and a codon encoding the N-terminal cysteine predicted to be the lipidation site highlighted in yellow.

2. The sequence of the new forward primer (highlighted in red) used to amplify the 5'-truncated *cchF* gene. This primer was used to mutate the cysteine codon (TGC) to methionine (ATG). The primer also included the first 15 bases immediately after the cysteine codon plus CACCATG on the 5' end of the primer, where CACC is required for insertion into the pET151/D-TOPO plasmid.

Figure 3-11: Design of the primers for the amplification of the 5'-truncated *cchF* gene.

The 1 kbps band was excised under UV light for subsequent purification using a gel purification kit. Subsequent purification of PCR product, ligation, transformation, selection of clones and PCR screening were performed analogously as with the full length *desE*, *cchF* and *cdtB*. The results of the PCR screening are described in Figure 3-13.

Table 3-4: The size of amplified DNA of the truncation mutants *desE, cchF* and *cdtB*

Gene	Size (DNA base pairs)
desE	962
cchF	967
cdtB	923

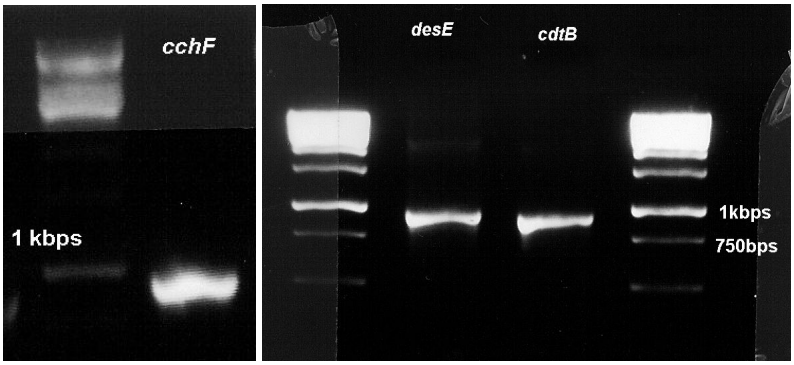


Figure 3-12: Analysis by agarose gel electrophoresis of the amplification of the 5'-truncated genes *cchF* (left), *desE* and *cdtB* (right), whose size was approximately 1 kbps.

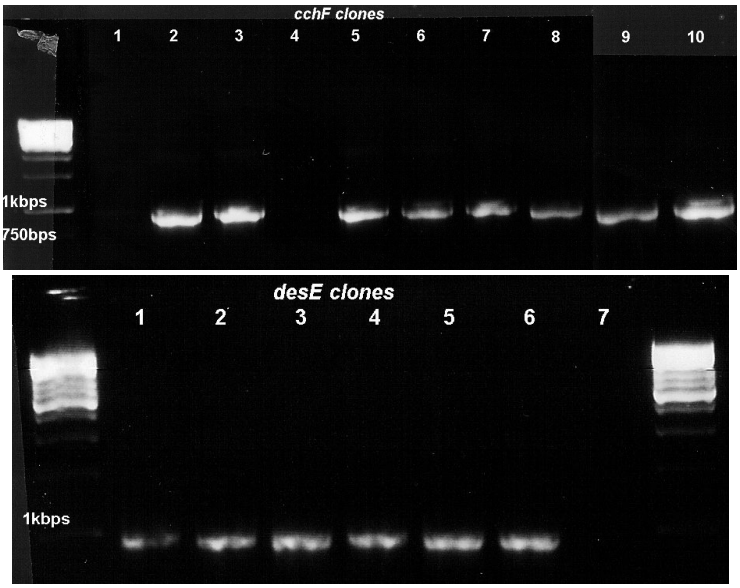


Figure 3-13: Agarose gel electrophoresis of the PCR screening of plasmid clones containing the 5'-truncated genes *desE* and *cchF* genes.

Minipreps of single colonies of *E. coli* TOP10 transformed with the products of the TOPO reaction pET151/D-TOPO + *cchF* (top) and pET151/D-TOPO + *desE* (bottom) were analysed by PCR with the same primers used in gene amplification to determine if the insert was present. In the case of pET151/D-TOPO + *cchF*, the clones 1 and 4 did not give a band at 1 kbps indicative of unsuccessful amplification of *cchF*. In the case of pET151/D-TOPO + *desE* clone 7 was rejected, because one could not be sure the insert would encode the correct protein.

3.4.4 : Overproduction of truncated His_6-DesE, His_6-CchF and His_6-CdtB proteins

E. coli BL21star was transformed with the appropriate plasmids and the truncated proteins overproduced as described previously for the full length constructs, and the proteins of the total cell extracts and cell-free extracts were analysed by SDS-PAGE. All three proteins were overproduced in the soluble fraction, and appeared as one large band (see Figure 3-14 and Figure 3-15) at molecular weights from 38 – 45 kDa (see Table 3-6). A few of the plasmids of overproducing clones were submitted for sequence analysis of their gene inserts. The pET151/D-TOPO plasmids containing the 5'-truncated amplimers *desE*, *cchF* and *cdtB* without mutations are denoted by the names given in Table 3-5. These plasmids were used to transform competent *E. coli* BL21star prior to large-scale protein overproduction.

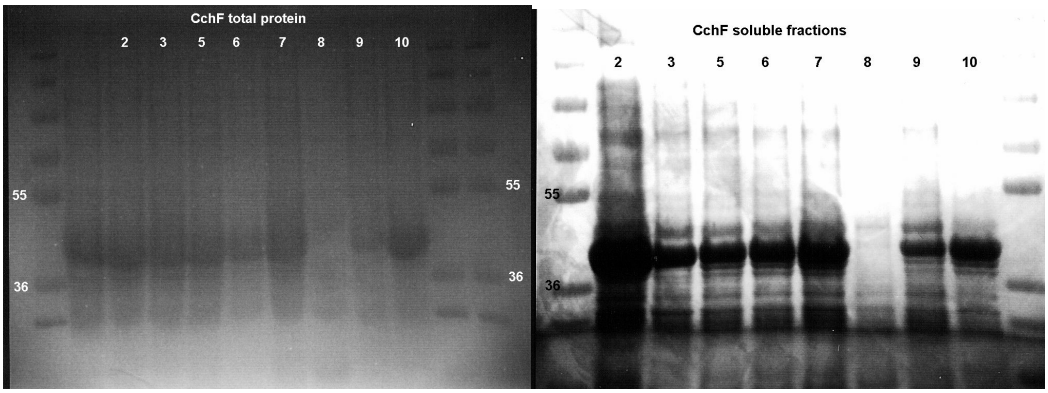


Figure 3-14: Analysis of the overproduction of the 5'-truncated protein His_6-CchF by different clones.

Left: The total protein fractions of truncated His_6-CchF overproducing clones of *E. coli* BL21star. Right: The soluble protein fraction (cell-free extract) of His_6-CchF overproducing clones of BL21star. Out of the clones tested for overexpression of a protein ~40 kDa, clone 8 did not overproduce a protein of the correct size.

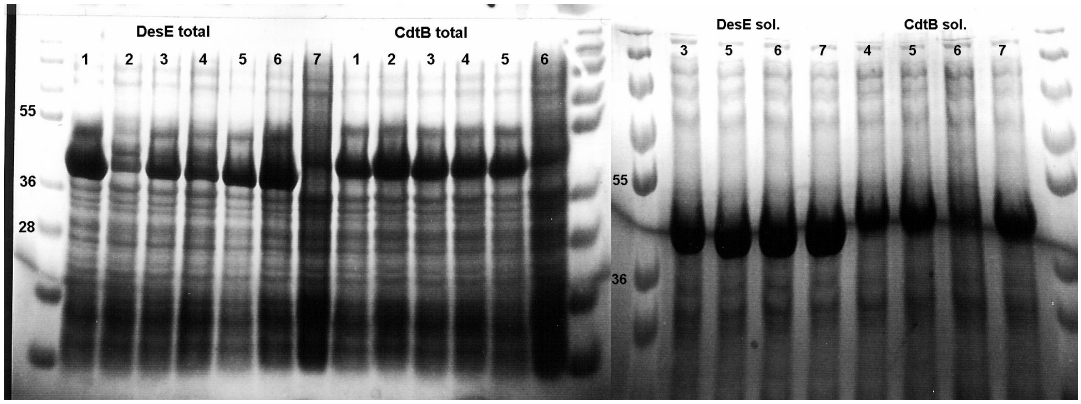


Figure 3-15: Analysis of overproduction of 5'-truncated His_6-DesE and His_6-CdtB proteins by clones by SDS-PAGE.

Left: The total protein content of truncated His_6-DesE and His_6-CdtB overproducing clones of *E. coli* BL21star. Right: The soluble protein fraction (cell-free extract) of truncated His_6-DesE and His_6-CdtB overproducing *E. coli* BL21star clones.

3.4.5 : Large-scale purification of 5'-truncated His_6-DesE, His_6-CchF and His_6-CdtB by Ni-NTA IMAC

This was performed using the procedure as described for the full length proteins His_6-DesE, His_6-CchF and His_6-CdtB. SDS-PAGE analysis of the IMAC purified proteins yielded a single band at around 40 kDa for all the proteins.

Table 3-5: Names of pET151/D-TOPO plasmids containing the 5'-truncated amplimers *desE*, *cchF* and *cdtB*.

Plasmid name	Gene inserted into pET151/D-TOPO plasmid
pPP004	5'-truncated *desE*
pPP005	5'-truncated *cchF*
pPP006	5'-truncated *cdtB*

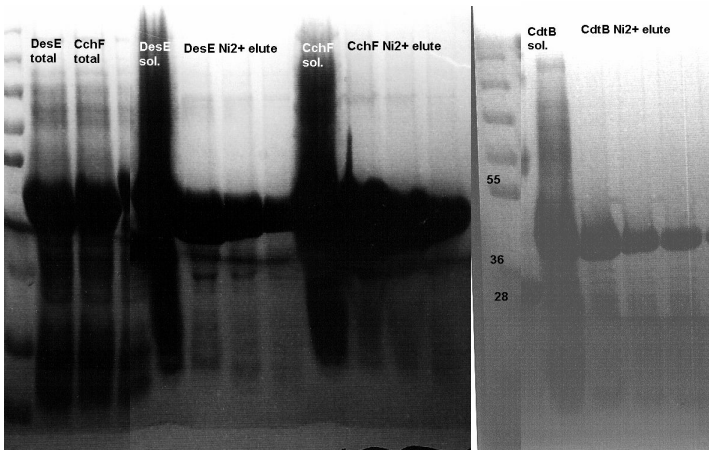


Figure 3-16: SDS-PAGE analysis of the purification of 5'-truncated His_6-DesE, His_6-CchF and His_6-CdtB proteins by Ni-NTA IMAC.

Left: the purification of His_6-DesE and His_6-CchF proteins, showing the total protein, soluble protein and elution by 300 mM imidazole from the Ni-NTA column, yielding a band of protein of the expected molecular weight. Right: the purification of His_6-CdtB protein, showing the soluble fractions and elution by 300mM imidazole from the Ni-NTA column, yielding a band of protein near the expected molecular weight.

3.4.6 : Proteolytic removal of N-terminal His_6 tags with TEV protease

The removal of the His-tag was performed to negate concerns that the tag may interact with transition metal ions and complexes in binding studies. Although it is assumed that His tags have little affinity for Fe(III), it would be difficult to rule out

whether His-tags interact with the transition metal were the cause of any unexpected phenomena that might be observed (e.g. fluorescence quenching by Fe(III) alone).

The cleavage of the N-terminal His-tags of the proteins was performed using TEV protease, the common name for the 27 kDa catalytic domain of the Nuclear Inclusion a (NIa) protein encoded by the tobacco etch virus[82]. TEV protease recognises the ENLYFQG motif in proteins and cleaves between the Q and G residues, which in this study results in the cleavage of the N-terminal histidine tag. This leaves an N-terminal tag of six residues (GIDPFT) followed by the methionine residue (mutated from cysteine) and the rest of the protein.

The recombinant TEV protease used contained a polyhistidine tag; thus to separate the proteolysed protein from the protease and undigested proteins, IMAC was used but instead the flow-through was collected, which should contain only the cleaved protein of the interest, with undigested proteins, cleaved His-tags and TEV protease remaining on the column.

TEV protease was overproduced in *E. coli* BL21(DE3)-RIL transformed with the plasmid pRK793, and purification of the protein was carried out analogously to other His-tagged proteins.

Table 3-6: Expected molecular weights of overproduced truncated His_6-DesE, His_6-CchF and His_6-CdtB before and after TEV protease digestion to remove the N-terminal His_6 tag.

Protein	Size (Da)	Size after TEV protease digest (Da)
His_6-DesE	37742.9	34590.4
His_6-CchF	38068.1	34915.6
His_6-CdtB	35922.8	32770.3

3.4.7 : Cleavage and purification of TEV protease digested proteins

Purified recombinant TEV protease was added to the His-tagged DesE, CchF and CdtB proteins to an approximate ratio of 1:50 and incubated at 4 °C overnight to keep the recombinant proteins cool. The digested protein was passed though a Ni-

NTA column and the flow-through collected and analysed by SDS-PAGE (Figure 3-17). The protein was concentrated and stored at -80 °C. The concentrations of purified proteins were determined by Bradford assay.

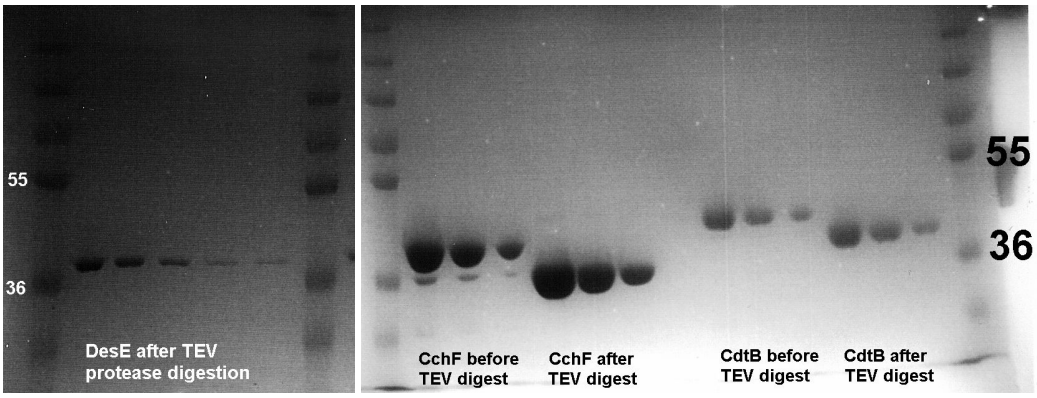


Figure 3-17: SDS-PAGE analysis of the 5'-truncated proteins His_6-DesE, His_6-CchF and His_6-CdtB after TEV protease digestion after purification via IMAC.

Left: Different concentrations of His_6- DesE after digestion with TEV protease. Right: Proteins His_6-CchF and His_6-CdtB before and after digestion with TEV protease. The apparent molecular weight decrease was indicative of successful TEV protease digestion.

3.5 : Characterization of purified, His-tag cleaved, 5'-truncated proteins DesE, CchF and CdtB

3.5.1 : Peptide mass fingerprinting of proteins

To confirm the identity of the truncated, TEV protease-digested proteins DesE, CchF and CdtB, peptide mass fingerprinting was carried out. In this procedure, electrospray ionization mass spectrometry (ESI-MS) was performed on tryptic digests of the proteins extracted from an SDS-PAGE gel. Matches between experimentally observed and theoretically predicted masses of fragments of peptides were observed: for truncated DesE, 15 peptide fragments were identified, giving 54% sequence coverage; for truncated CchF, 8 peptide fragments were identified, giving 32% sequence coverage and for truncated CdtB, 3 peptide fragments were identified, giving 16% sequence coverage.

3.5.2 : ESI-MS of proteins

The identity of the proteins was confirmed further using ESI-TOF-MS. The proteins were dialysed in 10 mM ammonium acetate buffer (pH 7.0) to remove traces of glycerol, to give a final protein concentration of 25 μM. The spectra of the proteins consisted of multiple peaks, corresponding to different multiply-charged states of a single species. The mass spectrum was deconvoluted to provide the mass-to-charge (m/z) ratio of the singly-charged ion.

The spectrum of truncated CchF consisted of several peaks (Figure 3-18) at m/z 3492.5244, 3175.0384, 2910.4381, 2686.7470 and 2494.9563, which corresponded to 10+, 11+, 12+, 13+ and 14+ charged ions. The spectrum was deconvoluted, and the m/z for the singly charged ion was deduced to be 34915 (Figure 3-19), which matches very well with the calculated m/z from its amino acid composition (34917). Similarly, the spectra of the truncated proteins DesE (Figure 3-20) and CdtB (Figure 3-21) were also deconvoluted by the software, to give the m/z values of the singly charged species of 34595 m/z and 32774 m/z respectively. These were very close to the theoretical m/z of the singly charged ions of DesE and CdtB (m/z 34591 and m/z 32771 respectively). These results further confirmed the identity of the recombinant proteins.

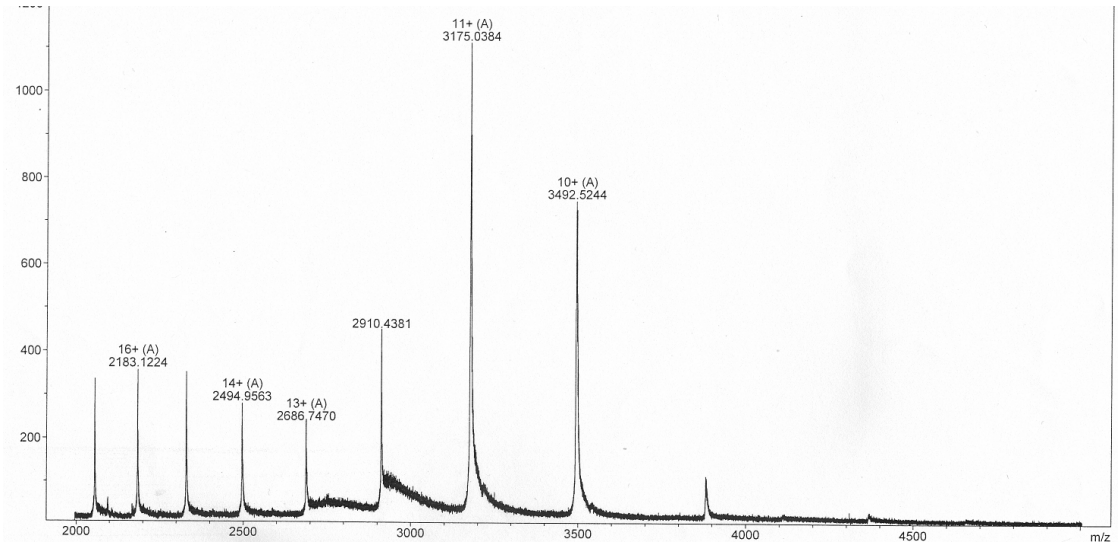


Figure 3-18: ESI-MS spectrum of truncated CchF (25 μM) in 10 mM ammonium carbonate buffer.

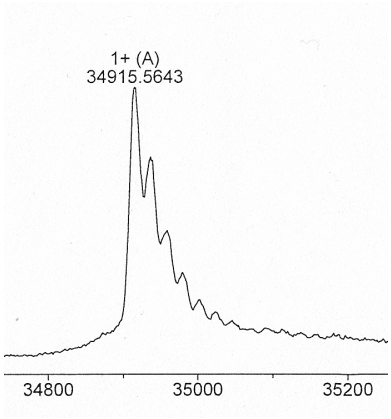


Figure 3-19: Deconvolution of ESI-MS spectrum of truncated CchF, giving the m/z of the singly charged ion.

The extra peaks are sodium adducts due to the insufficient acidification of the protein prior to injection into the ESI-MS.

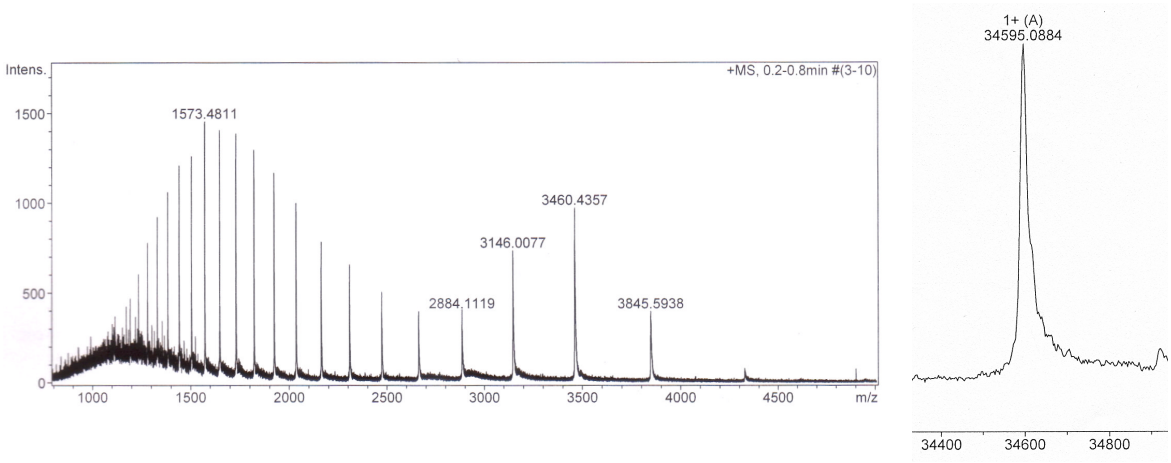


Figure 3-20: ESI-MS spectrum of truncated DesE and deconvolution.

Left: ESI-MS spectrum of truncated DesE (25 µM) in 10 mM ammonium carbonate buffer. The four peaks towards the right of the spectrum correspond to the 12+, 11+, 10+ and 9+ ions of DesE. Right: Deconvolution of ESI-MS spectrum of truncated DesE, showing mass of parent ion, giving the m/z for the singly charged ion.

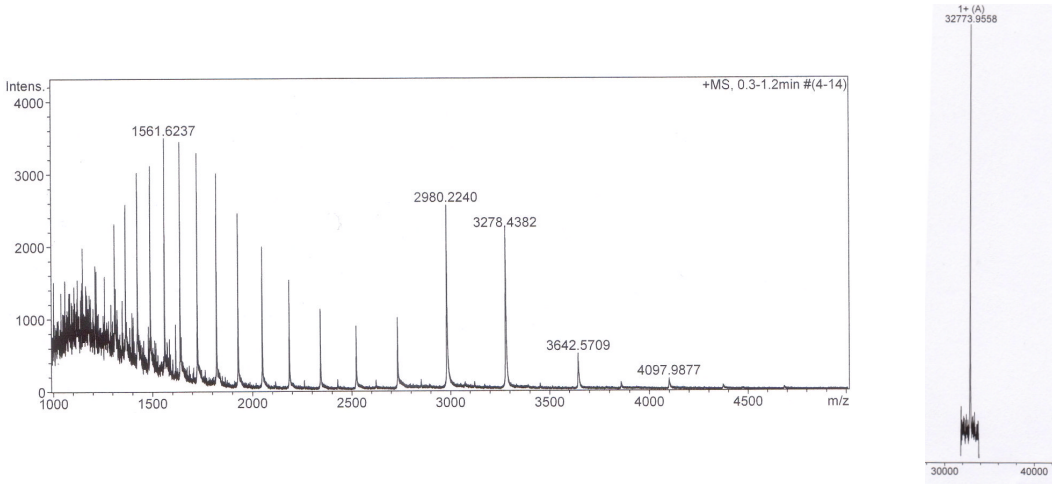


Figure 3-21: ESI-MS spectrum of truncated CdtB and deconvolution.

Left: ESI-MS spectrum of truncated CdtB (25 μM) in 10 mM ammonium carbonate buffer. The four peaks towards the right of the spectrum correspond to the 11+, 10+, 9+ and 8+ ions of the CdtB protein. Right: Deconvolution of ESI-MS spectrum of truncated CdtB, showing mass of parent ion, giving the m/z for the singly charged ion.

3.5.3 : Gel filtration analysis of truncated and His-tag cleaved DesE, CchF and CdtB

The proteins were analysed by gel filtration in the same way as the full length proteins. Fractions were collected and analysed by SDS-PAGE. DesE and CdtB eluted at volumes of 65.4 and 63.7 mL respectively (Figure 3-22 and Figure 3-23), which correspond to approximately 36.5 kDa and 41.7 kDa respectively. This demonstrated the proteins were monomeric. The CdtB protein eluted at lower volumes than DesE, indicating that this protein has a larger hydrodynamic radius than the latter, despite the lower molecular mass of CdtB.

CchF (Figure 3-24, top), eluted in two peaks at 56.9 and 63.6 mL respectively, which correspond to approximately 108 kDa and 42.4 kDa respectively. This implied the protein was a mixture of trimeric and monomeric forms. To obtain a monomeric form of the protein, an aliquot of protein CchF was incubated at 37 °C for 45 minutes. When analyzed by gel filtration (Figure 3-24, bottom), a large single peak eluted at 63.6 mL, indicating the protein was exclusively the monomeric form. This process was used prior to further experiments with truncated CchF to ensure work was carried out with a homogenous sample of protein.

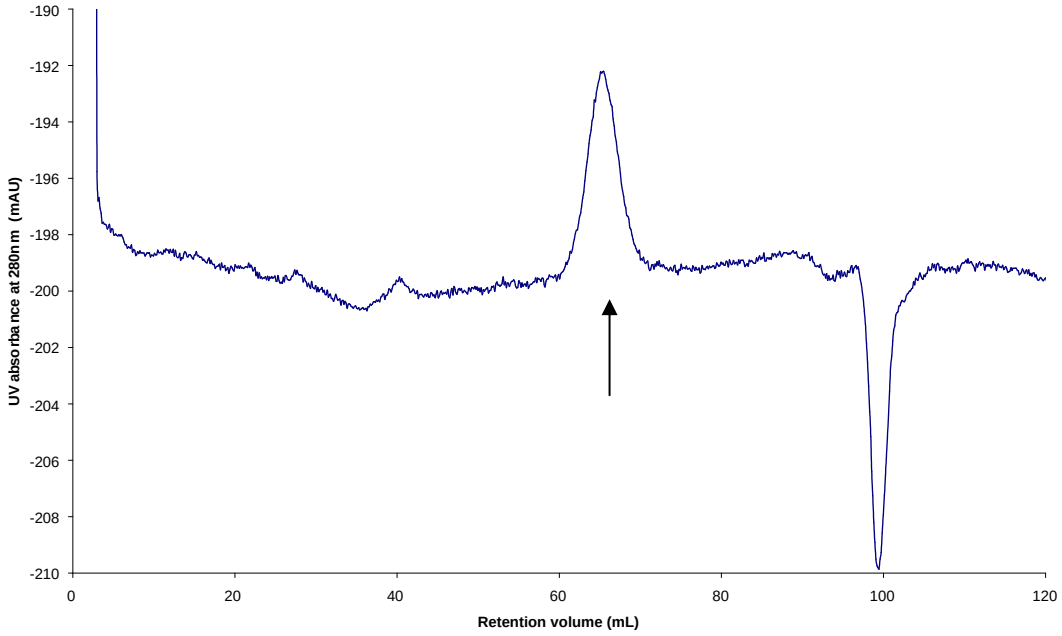


Figure 3-22: Gel filtration chromatograph of truncated DesE.

The protein eluted at 65.4 mL (arrow) which corresponds to a monomer. The dead volume of the column was 40 mL

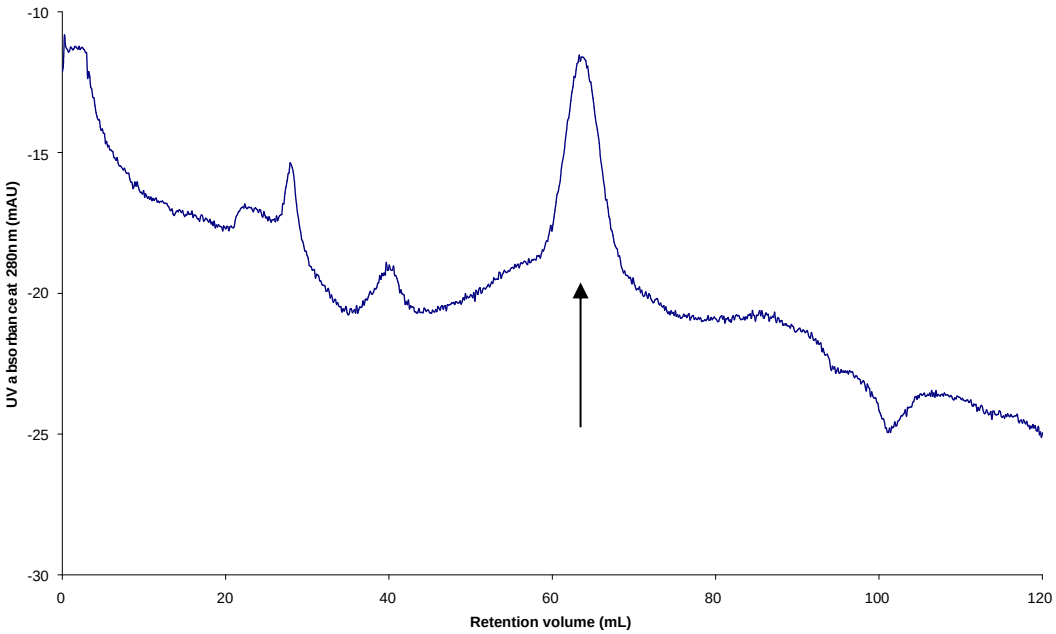


Figure 3-23: Gel filtration chromatograph of truncated CdtB.

The large peak (63.8 mL; arrow) corresponds to the monomeric form of CdtB. The other large peak at 40 mL is the dead volume of the column.

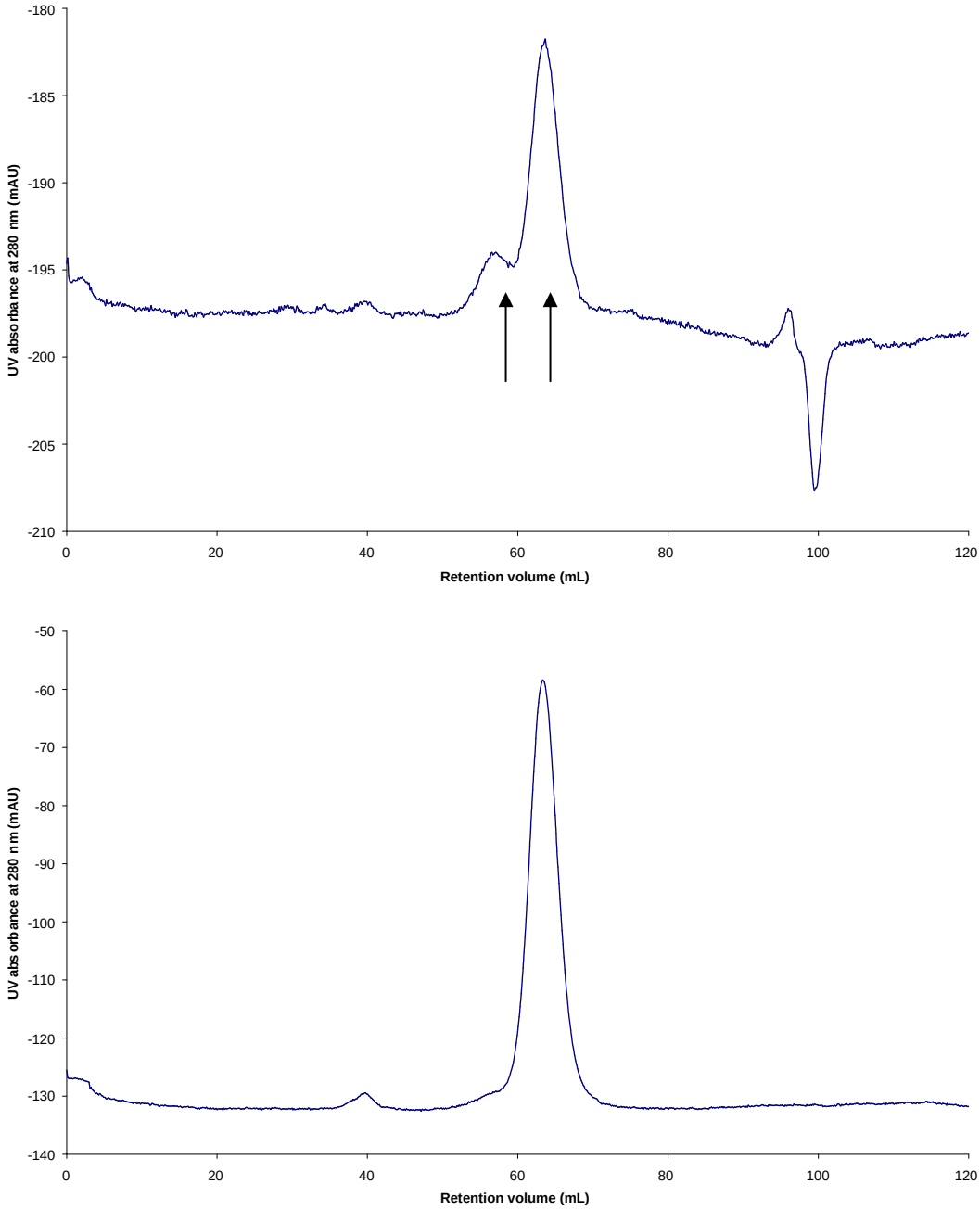


Figure 3-24: Gel filtration chromatographs of truncated CchF (A) without incubation (B) incubated at 37 °C for 45 minutes.

SDS-PAGE analysis of fractions corresponding to both peaks (indicated by arrows) of (A) corresponded to the same band, but analysis of their elution volumes (56.0 mL and 63.8 mL) corresponded to the former having a molecular weight three times that of the latter. Incubation of CchF at 37°C (B) eluted mostly the monomeric peak.

3.6 : Conclusions

The three *S. coelicolor* A3(2) lipoproteins DesE, CchF and CdtB were recombinantly overproduced in *E. coli* in a soluble form with N-terminal His-tags to enable purification via IMAC. Initially, the proteins were expressed with their N-terminal signal sequences, but these caused purification problems. When these were removed from the recombinantly overproduced proteins, along with mutation of N-terminal cysteine thought to be involved in lipidation to methionine, the truncated DesE, CchF and CdtB proteins were purified, in a mainly monomeric form.

The truncated proteins had their N-terminal His-tags cleaved by TEV protease, and the resulting proteins were characterised by peptide mass fingerprinting, mass spectrometry and gel filtration.

Chapter 4: Binding studies of receptor proteins with siderophores

4.1 : Intrinsic fluorescence quenching assay

4.1.1 : Introduction

Fluorescence generally is much more sensitive to the environment of a chromophore than is light absorption, hence is a most effective technique for following the binding of ligands to, or conformational changes in proteins[83].

Many proteins absorb strongly at ultraviolet wavelengths, and absorbance at 280 nm is mainly due to the aromatic residues tryptophan, tyrosine and to a much lesser extent phenylalanine[83]. These residues are also fluorescent in the ultraviolet, and in most proteins tryptophan fluorescence dominates[83], especially between the wavelengths 320 – 350 nm. Although tyrosine residues are usually present in larger numbers in proteins, and tyrosine is a slightly weaker emitter than tryptophan (emission wavelength 306 nm), its fluorescence is quenched by nearby tryptophans[83].

Also the maximum wavelength of tryptophan fluorescence is dependent on its environment: free L-tryptophan in water fluoresces at 348 nm, whilst tryptophan residues in hydrophobic environments can fluoresce as low as 320 nm, and this behaviour is exploited in structural, stability and binding studies, such as the gradual unfolding of proteins at high temperatures, where there is a red shift of the intrinsic fluorescence due to tryptophan residues being exposed to more polar environments.

One of the main uses of monitoring the fluorescence of tryptophan residues in proteins is the quenching of their fluorescence. Collisional quenching by the addition of O_2 or I^- ions in millimolar concentrations can give information about the exposure of tryptophan residues to the solution[83]. The other use exploited in this study is the quenching of tryptophan fluorescence caused by substrate binding[84], where monitoring the fluorescence after the titration of a substrate into a protein solution one can calculate rate constants and dissociation constants.

4.1.2 : Other considerations

The purity of the protein solution is important, as contaminants such as other proteins, DNA and particulates can interfere with the fluorescence measurements. Filtering of buffer solutions can remove many of the particulates.

One of the problems with measuring fluorescence is photobleaching caused by the lamp in the spectrometer, reducing the observed fluorescence after exposure to radiation. This may interfere with subsequent calculations if the titration of ligand with protein takes a large amount of time. The fluorimeter used in this study, a Perkin Elmer LS 50B, has a xenon flash lamp which emits a pulse of light at a rate of 50 Hz, which caused little photobleaching to a sample of tryptophan in water, hence this effect was not considered.

4.1.3 : Intrinsic fluorescence quenching with siderophore binding proteins

Intrinsic fluorescence quenching has been used in several studies to investigate the binding of ferri-siderophores to purified recombinant siderophore binding proteins[30,32,34,38,61]. In the case of ferric siderophores, the fluorescence of residues such as tryptophan can be quenched by the strong absorbance of the ferri-siderophore complex around the wavelength of tryptophan fluorescence (and most have little or no intrinsic fluorescence at this wavelength), hence ferric hydroxamates which are bound or very near to the protein can absorb the fluorescence. This phenomenon has been used to monitor the binding of ferric-hydroxamates and ferric-catecholates to proteins and dissociation constants were calculated on the assumption of a one-to-one protein-siderophore complex. In support of this assumption, some of these proteins have been co-crystallized with their substrates[28], and the structures of the apo forms of the proteins show only one binding cleft. All three proteins examined in this study (DesE, CchF and CdtB) contain tryptophan and tyrosine residues whose fluorescence could be quenched by addition of ferri-siderophores.

In these studies, there were common procedures that were followed. The temperature of the reaction used was 21 °C, and buffers were either 10 mM sodium phosphate buffer at pH 7.5, and were filtered prior to use. The protein was dissolved in the buffers to various concentrations (80 μg/mL – 1.6 μg/mL), ferri- and desferri-

siderophores were added and left to equilibrate for 4 minutes before making a fluorescence measurement.

The main assumption used in the analysis of intrinsic fluorescence quenching measurements is that the measure of fluorescence value results from a linear combination of the unbound and the bound protein, where the bound species has a lower fluorescence than the unbound protein. For example, if the fluorescence of the unbound species and bound species is 200 and 100 units respectively at 335 nm, and a solution of protein and ligand has a fluorescence of 150 units, then it is assumed that 50% of the protein is bound to the ligand in this solution. The linear transformation can be summarized as:

$$\text{Proportion of protein bound to ligand} = \frac{F_0 - F}{F_0 - F_{\min}} \quad \ldots (1)$$

where F_0 is the fluorescence of unbound protein, F is the fluorescence of a mixture of bound and unbound protein and F_{min} is the fluorescence of the fully bound protein.

Assuming the protein-ligand binding relationship is one to one, i.e. P + L ↔ PL then K_D can be calculated by fitting the proportion bound to a relation which is a function of $[P_T]$ and $[L_T]$, the total concentration of protein and ligand respectively:

$$\frac{F_0 - F}{F_0 - F_{\min}} = \frac{K_D + [P_T] + [L_T] - \sqrt{((K_D + [P_T] + [L_T])^2 - 4[P_T][L_T])}}{2[P_T]} \ldots (2)$$

This equation is derived directly from the definition of the dissociation constant $K_D = \frac{[P][L]}{[PL]}$ whose derivation is shown in Appendix A.2. In the previous studies mentioned, the raw fluorescence data was transformed by Equation 1, and then using a curve-fitting procedure a K_D was determined which produced curves (via Equation 2) that simulated the transformed data best.

4.1.4 : Why not perform a Scatchard analysis?

Scatchard analysis (the plot of $[L]_{free}/[L]_{bound}$ against $[P]_{bound}$ whose gradient is K_D) is a powerful way of calculating dissociation constants and the multiplicity of binding (n) of the $\mathbf{P} + n\mathbf{L} \leftrightarrow \mathbf{PL}_n$ type, but fitting data with this method does not put even bias on all data points (with points on the extremities of the dataset having undue influence on the results), unlike non-linear fitting methods where there is less bias.

4.1.5 : Estimating F_{min}

To estimate F_{min}, the fluorescence of the fully bound protein, the most straightforward way is to add a very high concentration of ligand, and in the case of ligands which are tight binders (i.e. low K_D) this is the most effective way. This may not be the true value of F_{min} as no amount of ligand is able to saturate the protein unless $K_D = 0$ μM, but the approximation would be adequate. For loose binders (i.e. high K_D) where not enough ligand can be added to saturate the protein sufficiently, and where the ligands have high absorbance of UV, such as ferri-siderophores, this method could be problematic as excess concentrations of these ligands could quench the fluorescence non-specifically.

Previous studies used a variety of methods[30,38] to estimate the value of F_{min}, where the derivation of their methods were not shown. More detailed analysis of these methods (see A.1) on two data sets whose F_{min} and K_D were known found that these methods did not estimate F_{min} accurately. Instead, the coupled equations as used to investigate tryptophan fluorescence quenching of human apolipoprotein M to retinol derivatives[85] were used, where the fluorescence quenching data was used to find both K_D and F_{min} simultaneously:

$$F = F_0\left(1 - \frac{[L]}{[L]+K_D}\right) + F_{min}\left(\frac{[L]}{[L]+K_D}\right)$$
$$[L] = \frac{[L_T]-K_D-[P_T]}{2} + \sqrt{\left(\frac{[L_T]-K_D-[P_T]}{2}\right)^2 + K_D[L_T]} \quad \ldots(3)$$

$[L]$ is the concentration of free ligand and is a variable which was used to couple the two equations. $[P_T]$ and $[L_T]$ are the total concentration of protein and ligand respectively. This set of equations was derived from the definition of $K_D = \frac{[P][L]}{[PL]}$, and reduces to Equation 2 if F_{min} was already known. This method was used to estimate F_{min} and K_D simultaneously if F_{min} cannot be estimated from addition of a high concentration of ligand.

4.2 : Siderophore concentration determination

4.2.1 : Preparation of siderophore solutions

All ferri-siderophores and desferri-siderophores were obtained from commercial suppliers, except for ferri-coelichelin which was purified from *S. coelicolor* as described in Chapter 2. Desferrioxamine B was purchased as its mesylate salt, whilst ferrioxamine B was produced by the reaction of desferrioxamine B with an excess of ferric chloride, and was purified by RP-HPLC prior to experiments. Ferri-siderophore and desferri-siderophore solutions were prepared by dissolving in deionized H_2O.

4.2.2 : Spectrophotometric concentration determination of ferri-siderophore solutions

Ferric-siderophore complexes are usually highly coloured, for example ferrioxamine B is red and ferri-enterobactin is a dark violet. The red-orange-yellow colour of ferric hydroxamate solutions in particular is due to the high absorbance of ferric *tris* hydroxamates at approximately 435 nm. This is due to the metal-to-ligand charge transfer transition from the Fe^{3+} ion to the hydroxamate ligands. The concentrations of ferric-siderophore stock solutions can be determined by using published extinction coefficients[11], and measuring absorbance at around 435 nm. After obtaining the concentrations of the stock solutions of the ferri-siderophores, they were diluted with ultrapure water to produce 500 μM, 100 μM and 10 μM stock solutions.

4.2.3 : ICP-OES of ferri-siderophore solutions

The concentration of iron in the 500 μM and 100 μM solutions of the ferric complexes were also measured by inductively coupled plasma optical emission spectroscopy (ICP-OES). Because there is one iron atom in each ferri-siderophore complex, the concentration of iron in the solution directly corresponds to the concentration of siderophore (assuming there is negligible free iron in the solution). The concentrations determined by ICP-OES corresponded well to the concentrations estimated by absorbance, except in the case of ferri-albomycin, where ICP-OES indicated the stock solutions were half the concentration as estimated from absorbance. This was an unexpected result, which may suggest that ferri-albomycin forms a 2 ligand – 1 iron complex or the extinction coefficient used was inaccurate. The latter was assumed, as crystal structures of ferri-albomycin bound to siderophore-binding proteins show a one-to-one ligand-iron complex[23,28].

As the concentrations of the different ferri-siderophore solutions calculated from ICP-OES can be directly compared against each other, these values were used for all the experiments described below rather than the concentrations obtained by absorbance, except for ferri-enterobactin and the desferri-siderophore solutions, whose concentrations were calculated from the dilution of stock solutions of concentration approximately 1 mg/mL.

Table 4-1: Concentration of ferri-siderophore solutions (approximately 500 μM and 100 μM as calculated by absorbance) as estimated by ICP-OES

Siderophore solution	Concentration estimated by ICP-OES (μM)	
	~500 μM	~100 μM
Ferrichrome	550 ± 10	110 ± 3
Ferrioxamine B	750 ± 10	150 ± 3
Ferrioxamine E	450 ± 10	90 ± 3
Ferri-albomycin	250 ± 7	50 ± 2
	~300 μM	
Ferri-coelichelin	300 ± 25	

4.3 : Intrinsic fluorescence quenching of DesE

4.3.1 : Fluorescence spectra of DesE in the absence and presence of ferri-siderophores

Purified DesE (5'-truncated, His_6 tag cleaved) was dissolved in phosphate-buffered saline solution (PBS) to total concentration 78 µg/mL, which had been filtered through 0.22 µm cellulose acetate filters prior to use. The instrument used in the fluorescence quenching experiments was a Perkin Elmer LS 50B luminescence spectrometer with excitation wavelength of 280 nm with excitation and emission slit lengths of 5 nm and 8 nm respectively. The fluorescence spectrum was recorded from 310 nm to 380 nm.

The spectrum of DesE without siderophores had a maximum at 337 nm. On addition of small concentrations of the ferri-siderophores ferrioxamine B, ferrioxamine E, ferrichrome and ferri-albomycin, a large decrease of fluorescence was observed (Figure 4-1). This effect was not observed on addition of similar concentrations of ferri-coelichelin and ferri-enterobactin. This was also the case on addition of the desferri-siderophores desferrioxamines E and B, desferrichrome and enterobactin.

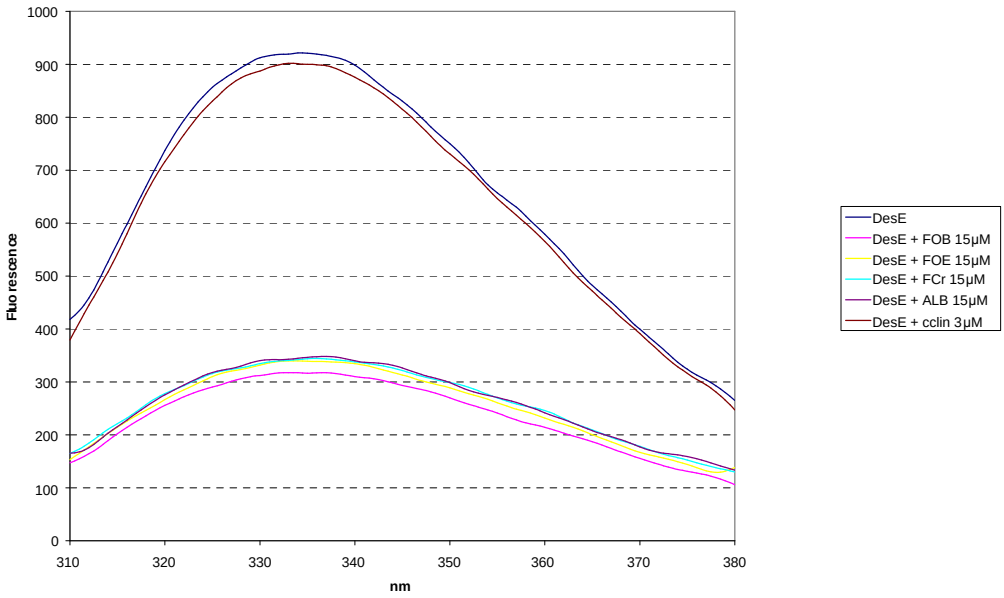


Figure 4-1: Fluorescence spectra of DesE with and without of ferri-siderophores

Ferri-siderophores added: ferrioxamine B (FOB), ferrioxamine E (FOE), ferrichrome (FCr), ferri-albomycin (ALB) and ferri-coelichelin (cclin).

Based on the precedent of previous studies[30,38,61], this was almost certainly due to the specific binding of DesE to the ferrioxamines, ferrichrome and ferri-albomycin, but not to ferri-coelichelin. At 335 nm, the fluorescence of DesE decreased up to 65% on addition of the ferri-siderophores (Figure 4-2), with nanomolar concentrations of siderophore able to quench the fluorescence significantly.

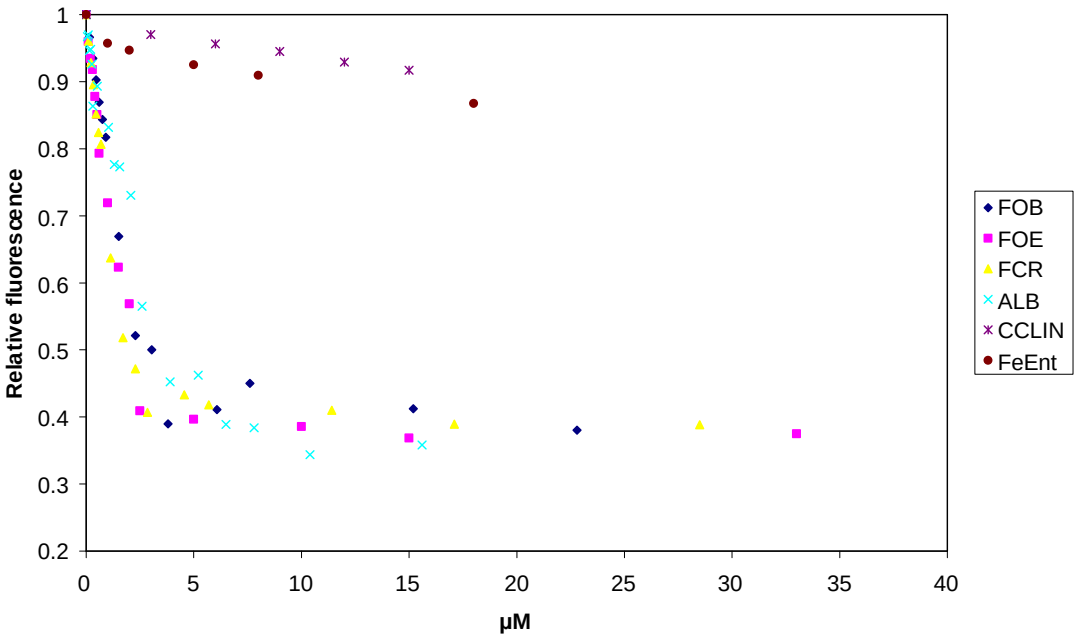


Figure 4-2: Relative decrease of the fluorescence at 335 nm of DesE on addition of ferri-siderophore.

Key: ferrioxamine B (FOB), ferrioxamine E (FOE), ferrichrome (FCR), ferri-albomycin (ALB), ferri-coelichelin (CCLIN) and ferri-enterobactin (FeEnt).

4.3.2 : Estimation of dissociation constants

Due to the low amount of DesE fluorescence quenching caused by ferri-coelichelin, ferri-enterobactin and the desferri-siderophores, a dissociation constant was not determined for these ligands. To calculate dissociation constants for the other ferri-siderophores, the fluorescence was monitored at 335 nm without siderophores (F_0) and with various concentrations of siderophore (F). The protein solution in the fluorescence cuvette had a volume of 90 µL and no more than 4.5 µL of siderophore solution was added to this solution. The effect of dilution of the protein on the

fluorescence of the protein solution was assumed to be negligible as the same volume of pure water reduced the fluorescence on average less than 3 %.

F_{min} was estimated by addition of a high concentration of siderophore as DesE was saturated easily (the fluorescence did not change very much on addition of concentrations of ferri-siderophore higher than 5 μM). Equation 1 was used to calculate the proportion of protein bound to the ligand, using Microsoft Excel. Fluorescence values from the average of at least three separate titrations were used to calculate the proportion of bound protein as a function of ligand concentration. The data was subsequently entered into the data analysis program Origin 6.0 Professional[86].

To find K_D, the proportion of bound protein was simulated using Equation 2, which expresses this as a function of K_D, the total protein concentration [P_T] and the total ligand concentration [L_T]. [P_T] was determined by using the molecular weight as calculated from the amino acid sequence of DesE (Table 4-2). As [L_T] was the only variable, only K_D was varied to fit the curve. The "Non-linear Curve Fit" function was used to find the best K_D, using the function file Onesitebind2 (code in Appendix A.3.1). The data points and curve fits for relative binding against ferri-siderophore concentration added are given in Figure 4-3.

Table 4-2: The molecular masses of DesE, CchF and CdtB and the molar concentration in a 78 μg/mL solution.

Protein	Molecular weight (g/mol)	Concentration of 78 μg/mL solution (μM)
DesE	34590	2.255
CchF	34916	2.234
CdtB	32770	2.380

The three ferri-siderophores ferrioxamine B, ferrioxamine E and ferrichrome bound to DesE with nanomolar affinity, whilst the protein had a weaker affinity to ferri-albomycin (Table 4-3). The dissociation constants were the same range for those obtained for other Gram-positive siderophore binding proteins[38,61]. Although the ferrichrome analogue ferri-albomycin bound less strongly to DesE than ferrichrome itself, the K_D indicates that ferri-albomycin is recognised by DesE. This is an

interesting result, as *S. coelicolor* is not sensitive to this antibiotic[59]. There are several explanations for this: the affinity of DesE to ferri-albomycin is simply too low to allow efficient transport of the antibiotic; the DesE-ferri-albomycin complex may not be recognised by the permease parts of the ABC transporter associated with DesE, or that it is imported by the ABC transporter, but the cell is able to inactivate the molecule or export it out the cell, e.g. via a multidrug transporter[87]

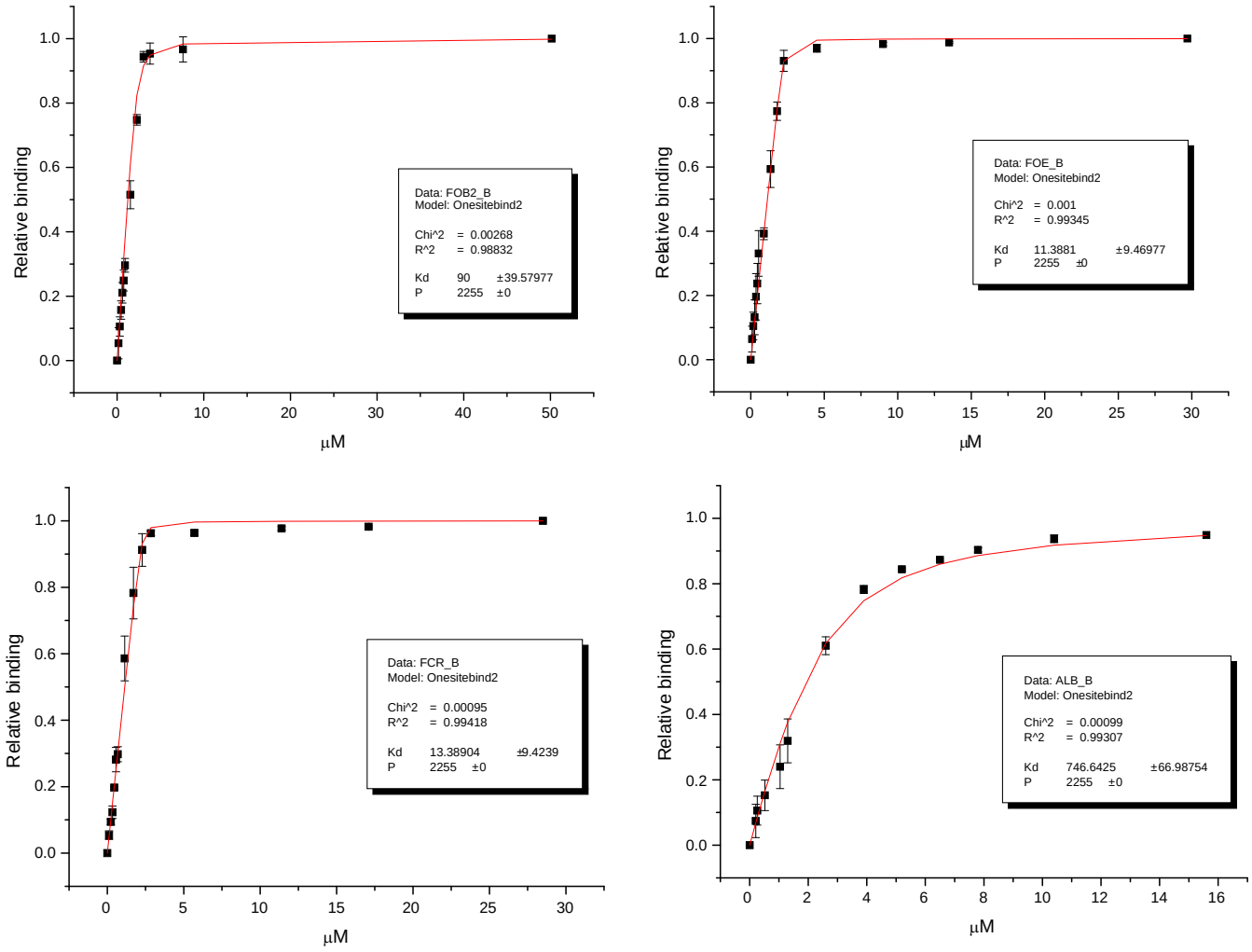


Figure 4-3: The relative binding of various ferri-siderophores to a DesE solution (78 μg/mL) at 21 °C.

(Top left) ferrioxamine B; (Top right) ferrioxamine E; (Bottom left) ferrichrome; (Bottom right) ferri-albomycin. The data was fitted to the Onesitebind2 model, and the non-linear curve fitting function was used to find a suitable K_D given [P] = 2.255 μM. The K_D values are given in nM.

The results of this experiment were consistent with the hypothesis of Barona-Gómez *et al.*[53], based on the genetic experiments that DesE can bind to ferrioxamine E and ferrichrome, but not ferri-coelichelin

Table 4-3: Calculated dissociation constants for ferri-siderophores from DesE. (N.D. – not determined as there was insufficient quenching to calculate a dissociation constant)

	K_D (μM)
Ferrioxamine B	0.090 ± 0.040
Ferrioxamine E	0.011 ± 0.009
Ferrichrome	0.013 ± 0.009
Ferri-coelichelin	N. D.
Ferri-albomycin	0.750 ± 0.066

4.4 : Intrinsic fluorescence quenching of CchF

4.4.1 : Fluorescence spectra of CchF in the absence and presence of ferri-siderophores

The fluorescence spectrum of CchF (5'-truncated, His_6 tag cleaved) was rather weak compared to DesE, hence the same spectrometer conditions were used as with DesE except the emission slit length was increased from 8 nm to 10 nm to obtain a more intense signal. Also CchF had an emission maximum at 325 nm, which may suggest the tryptophan residues that contribute most of the fluorescence at this wavelength are in hydrophobic environments. The fluorescence of CchF was monitored at 325 nm.

The addition of small amounts of ferri-coelichelin was able to quench the fluorescence of CchF significantly, indicating a specific binding of this siderophore. Relatively high concentrations of ferrioxamine B and ferri-albomycin were needed to quench the fluorescence of CchF significantly, indicating a weaker interaction between these siderophores and CchF. High concentrations of the ferri-siderophores ferrioxamine E and ferrichrome were unable to quench CchF sufficiently to allow dissociation constants to be determined. This was also the case with desferrioxamine B, E, desferri-ferrichrome, enterobactin and ferri-enterobactin.

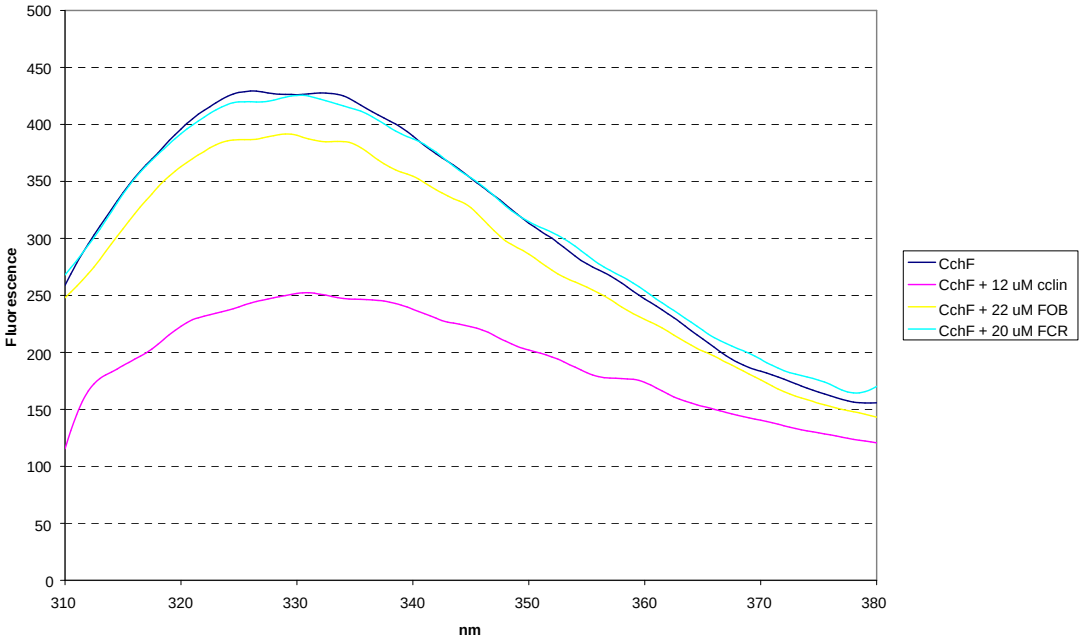


Figure 4-4: Fluorescence spectra of CchF with and without of ferri-siderophores

Ferri-siderophores added: ferri-coelichelin (cclin), ferrioxamine B (FOB) and ferrichrome (FCR).

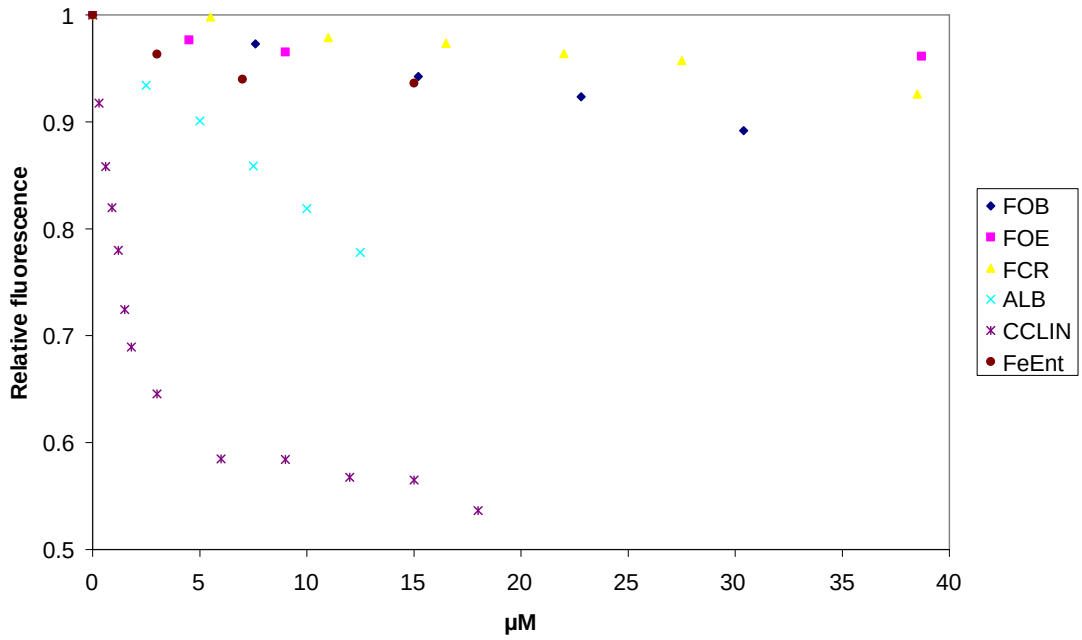


Figure 4-5: Relative decrease of the fluorescence at 325 nm of CchF on addition of ferri-siderophores.

Key: ferrioxamine B (FOB), ferrioxamine E (FOE), ferrichrome (FCR), ferri-albomycin (ALB), ferri-coelichelin (CCLIN) and ferri-enterobactin (FeEnt).

4.4.2 : Estimation of dissociation constants

To estimate dissociation constants, the fluorescence was monitored at 325 nm. F_{min} was estimated and K_D was found in the case of ferri-coelichelin using the methods that were described for DesE. The graph and curve fit of relative binding against ferri-coelichelin concentration is given in Figure 4-6 (top).

In the case of ferrioxamine B and ferri-albomycin, full saturation of CchF was not possible, hence F_{min} needed to be estimated. To find K_D and F_{min} simultaneously, Ahnström *et al.* used the coupled equations (Equation 3) to fit the raw fluorescence data, but for this experiment this data was unreliable as the F_0 of the protein solution was different from day-to-day; this would introduce different errors compared to the curve for binding of ferri-coelichelin to CchF and it would be unreasonable to compare the datasets.

Instead, the data was transformed by Equation 1 but using the F_{min} calculated for ferri-coelichelin, to give values of "relative fluorescence decrease compared to fully saturated CchF with ferri-coelichelin" using Microsoft Excel, and then the equations of Ahnström *et al.* were used to fit the data to obtain values of K_D and $Q = \frac{F_0 - F_{\min}}{F_0 - F_{\min}(coelichelin)}$, whose physical interpretation is the quenching of the protein by the ligand compared to ferri-coelichelin. For example, if Q = 2, then the ligand caused twice as much fluorescence quenching at maximal concentrations than ferri-coelichelin, whilst if Q = 0.5, the ligand caused only half as much fluorescence quenching at maximal concentrations than ferri-coelichelin. Optimizing the parameter Q is analogous to optimizing F_{min} directly.

Again, the program Origin 6.0 Professional was used to fit the transformed data for quenching of CchF fluorescence by ferrioxamine B and ferri-albomycin, and the "Non-linear curve fitting" function was used but instead the function file Onesitebind_lipid based on equation 3 from Ahnström *et al.*[85] (named after the Journal of Lipid Research where the paper was published: see Appendix A.3.2 for the function file) was used to optimise simultaneously both K_D and Q.

Q was optimized with K_D for the ferrioxamine B and ferri-albomycin datasets, with *Q* estimated to be 0.5 and 1.3 respectively. The physical interpretation of these values is that ferrioxamine B would only quench CchF half as much as ferri-coelichelin at saturating concentrations, whilst ferri-albomycin would quench 1.3 times as much. The data points and curve fit of "relative fluorescence decrease compared to F_{min} (*coelichelin*)" against ferrioxamine B and ferri-albomycin concentration is given in Figure 4-6.

It is clear from these results that ferri-coelichelin has a very specific interaction with CchF compared to the other ferri-siderophores tested. The dissociation constants calculated for ferrioxamine B and ferri-albomycin are probably too weak for CchF to mediate their uptake. This observation is consistent with the findings of Barona-Gómez *et al.*[53] that *S. coelicolor* mutants whose *des* cluster (including the *desE* gene) was removed but whose *cch* cluster remained grew poorly when the medium was supplemented with ferrioxamine B and ferrichrome, but grew well when the medium was supplemented with ferri-coelichelin.

Table 4-4: Calculated dissociation constants of ferri-siderophores from CchF (N.D. – not determined as there was insufficient quenching to calculate a dissociation constant)

	K_D (μM)
Ferrioxamine B	37.1 ± 6.4
Ferrioxamine E	N. D.
Ferrichrome	N. D.
Ferri-coelichelin	0.33 ± 0.02
Ferri-albomycin	20.4 ± 3.6

The data also support the notion that the *cch* cluster encodes a self-contained siderophore synthetase/uptake system where iron acquisition is dependent only on coelichelin production and uptake. This gene cluster could have evolved separately in *S. coelicolor* or in another unknown organism from which it could have been horizontally transferred to *S. coelicolor*[18].

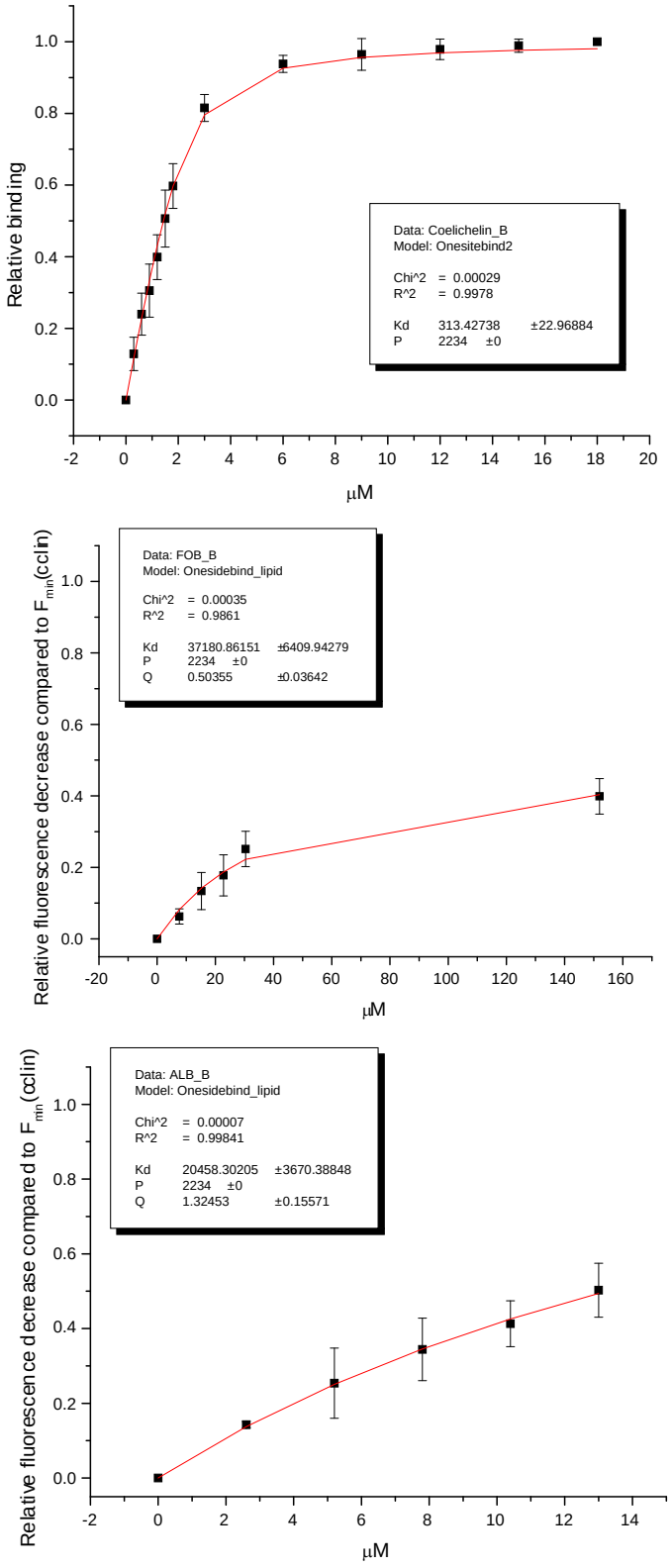


Figure 4-6: The relative binding of ferri-siderophores to a CchF solution (78 µg/mL) at 21 °C.

(Top) The relative binding of ferri-coelichelin to CchF. The data was fitted to the Onesitebind2 method, and the non-linear curve fitting function was used to find a suitable K_D given [*P*] = 2.234 µM. (Middle) The relative fluorescence decrease compared to F_{min}(*coelichelin*) for

ferrioxamine B binding to CchF. (Bottom) The relative fluorescence decrease compared to F_{min}(*coelichelin*) for ferri-albomycin binding to CchF. The datasets was fitted to the Onesitebind_lipid model, and the non-linear curve fitting function was used to find a suitable K_D and Q given [P] = 2.234 µM.

4.5 : Intrinsic fluorescence quenching of CdtB

4.5.1 : The decrease of the fluorescence of CdtB

The conditions and instrument parameters were identical to those described for DesE. CdtB (5'-truncated, His_6 tag cleaved) had a fluorescence maximum at 335 nm, hence changes in fluorescence intensity was monitored at this wavelength. Unlike DesE and CchF, the fluorescence of CdtB decreased with time (over a few hours) when left to stand in solution at 21 °C (Figure 4-7). Also, the wavelength of the fluorescence maximum of the spectrum increases from 335 nm to 340 nm (Figure 4-8). This is likely to be due to the instability of the protein at room temperature. The tryptophan residues which reside in relatively hydrophobic environments in the folded protein, are exposed to the hydrophilic solvent as the protein unfolds, resulting in quenching of their fluorescence and a red-shift in the emission towards the emission maximum of free L-tryptophan in water (348 nm).

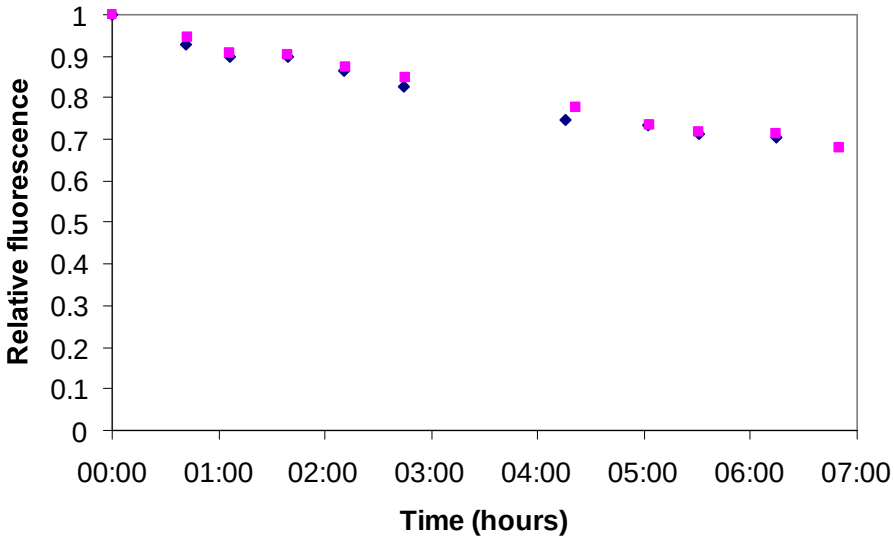


Figure 4-7: Relative fluorescence at 335 nm of CdtB (78 µg/mL) at 21 °C over the course of seven hours.

4.5.2 : Spectra of CdtB on addition of ferri-siderophores

Even after standing for seven hours at room temperature, the decrease in fluorescence of CdtB upon addition of ferri-siderophores was quite large (Figure 4-9), indicating there was still sufficient folded protein to examine binding by fluorescence quenching. Most of the experiments were carried out within three hours, and the protein was assumed to remain largely folded over this period.

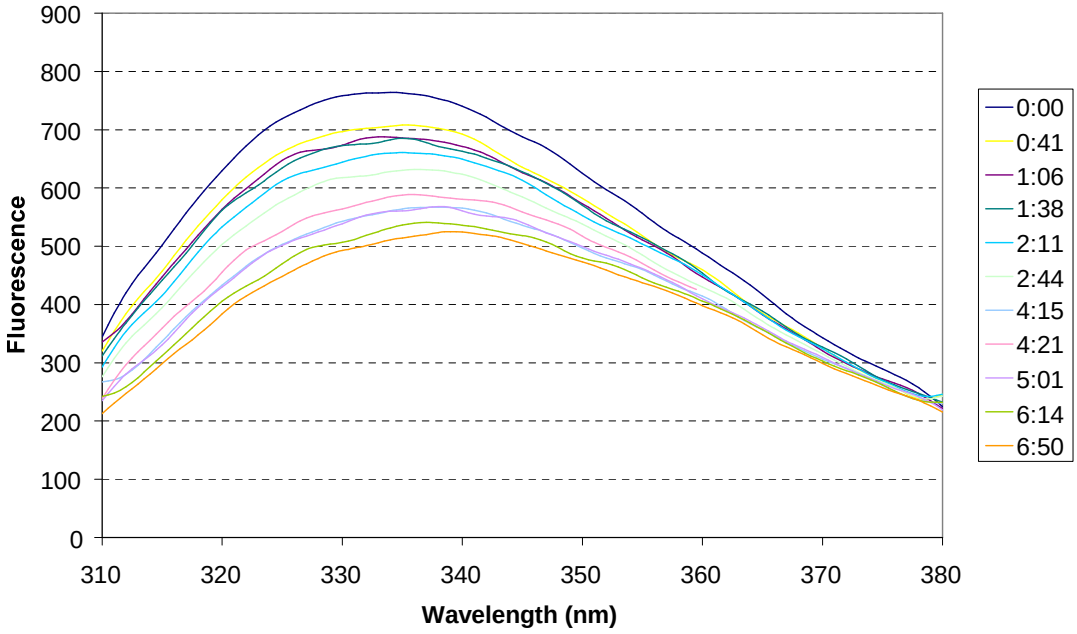


Figure 4-8: The fluorescence emission from 310 nm to 380 nm of CdtB (78 µg/mL) over 7 hours. The legend labels show the time elapsed (hours:minutes) after preparation of the protein solution.

Surprisingly, both ferrioxamine B and ferri-coelichelin were found to quench the fluorescence of CdtB strongly at low concentrations (Figure 4-9). Only ferrioxamine B has been implicated by genetic experiments to be a substrate for the ABC transporter system CdtABC[53]. Ferri-albomycin was found to quench the fluorescence of CdtB significantly but only at higher concentrations than ferrioxamine B and ferri-coelichelin. Ferrichrome and ferrioxamine E were less able to quench the fluorescence of CdtB, even at high concentrations. Ferri-enterobactin and the desferri-siderophores tested were unable to quench the fluorescence of CdtB at all.

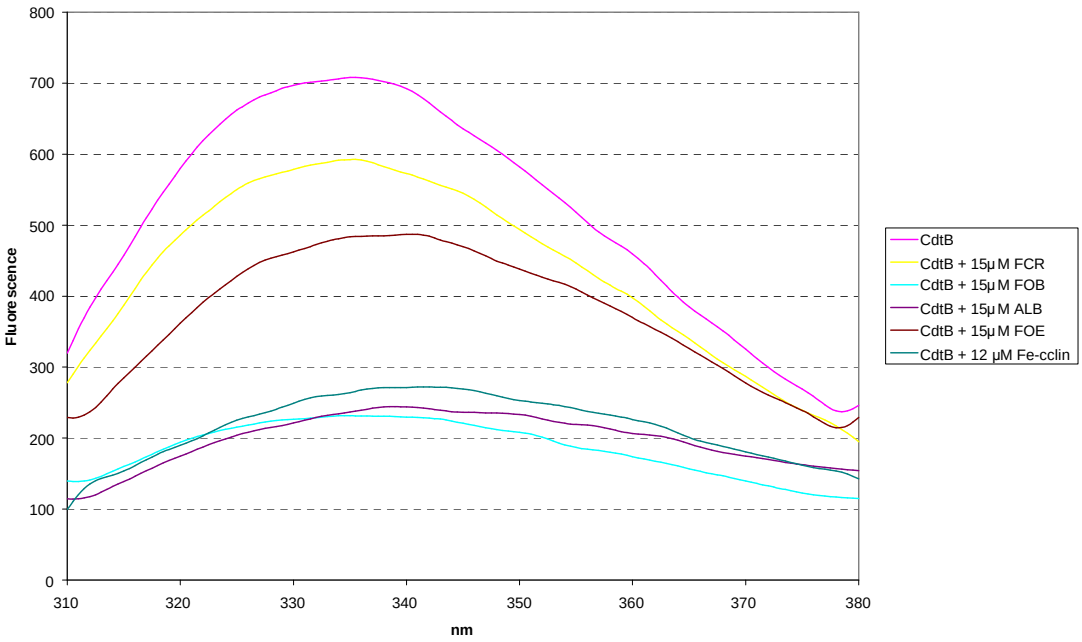


Figure 4-9: Fluorescence spectra of CdtB with and without of ferri-siderophores

Ferri-siderophores added: ferrichrome (FCR), ferrioxamine B (FOB), ferri-albomycin (ALB), ferrioxamine E (FOE) and ferri-coelichelin (cclin).

4.5.3 : Estimation of dissociation constants

The binding of ferri-siderophores to CdtB appears complicated (Figure 4-10). Ferrioxamine B and ferri-coelichelin are tight binders, whereas ferrichrome and ferrioxamine E are much weaker binders to CdtB. Both ferri-coelichelin and ferrioxamine B caused a similar decrease in the fluorescence of CdtB. F_{min} in the case of ferrioxamine B and ferri-coelichelin was determined to be the fluorescence of the protein at high concentrations of ferri-siderophore.

Because the protein was partially unfolded during the measurements, the actual value of [P] was difficult to estimate, but with no way of determining the proportion of folded to unfolded protein, the initial protein concentration was used, as well as 75% of this value (assuming that 25% of the protein was unfolded hence could not bind ferri-siderophores). This value will be explained in the next section.

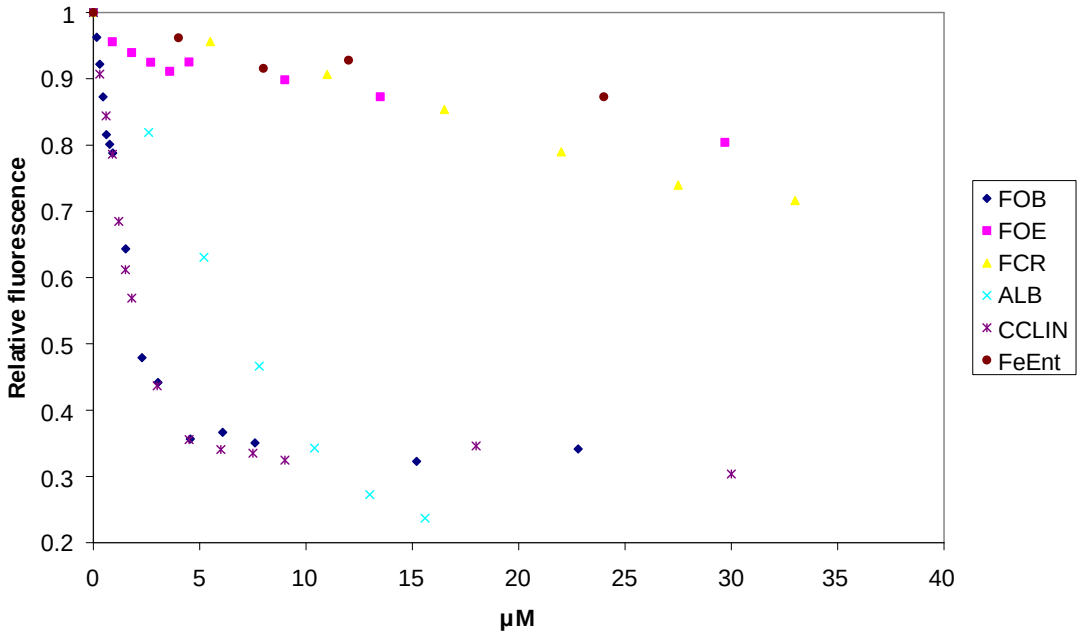


Figure 4-10: Relative decrease of the fluorescence at 335 nm of 78 mg/mL CdtB on addition of various amounts of siderophore.

Ferri-siderophores added: ferrioxamine B (FOB), ferrioxamine E (FOE), ferrichrome (FCR), ferri-albomycin (ALB), ferri-coelichelin (CCLIN) and ferri-enterobactin (FeEnt).

To estimate K_D for ferri-coelichelin and ferrioxamine B, where F_{min} was determined to be the residual fluorescence at high concentrations of ferri-siderophore, the proportion of protein bound as a function of ligand concentration was simulated by Equation 2. The "Non-linear Curve Fit" function was used to find the best K_D, using the function file Onesitebind2. The graphs and curve fits for relative binding against ferri-coelichelin and ferrioxamine B concentrations are given in Figure 4-11.

To estimate K_D in the case of ferri-albomycin, ferrioxamine E and ferrichrome, the F_{min} values used for ferrioxamine B binding were used because F_{min} could not be determined due to the high concentrations of ferri-siderophore needed to saturate the protein. Equation 1 was used to calculate the "relative fluorescence decrease compared to fully saturated CdtB with ferrioxamine B" in Microsoft Excel, and the "Non-linear Curve Fit" function in Origin 6.0 was used to fit the data to the model described by Ahnström *et al.*[85] (Onesitebind_lipid), where K_D and $Q = \frac{F_0 - F_{\min}}{F_0 - F_{\min}(FOB)}$ were optimized. Ferri-albomycin caused greater total quenching

than ferrioxamine B, hence *Q* had a large value (1.9), which was also the case with ferrichrome (1.3). Ferrioxamine E was unable to cause as much quenching of the fluorescence of CdtB as ferrioxamine B, even at very high concentrations (77 μM), but also the fluorescence decrease seemed to plateau at lower concentrations. The value of *Q* determined (0.43) can be interpreted as ferrioxamine E causing less than half the fluorescence decrease caused by ferrioxamine B.

From the values given in Table 4-5, both ferrioxamine B and ferri-coelichelin bind with high specificity to CdtB. The research carried out by Barona-Gómez *et al.*[53] showed that non-siderophore producing mutants of *S. coelicolor* lacking the complete *des* and *cch* clusters were able to grow supplemented with ferrioxamine B as an iron source, thus indicated that other siderophore uptake systems, such as the ABC transporter system encoded by the *cdtABC* gene cluster could be used in uptake of this siderophore.

S. coelicolor M145 is sensitive to the antibiotic ferri-salmycin[88], a ferrioxamine B analogue produced by the bacterium *Streptomyces violaceus* DSM 8286, but the *S. coelicolor* M145 mutant without the *cdtABC* gene cluster was resistant to the antibiotic[59]. Also it was shown that the sensitivity was reduced in *S. coelicolor* M145 on supplementation of ferrioxamine B[59], indicating the latter was interfering with transport of ferri-salmycin via the CdtABC system. It would have been desirable to confirm that the siderophore binding protein used to bind ferri-salmycin before transport into cells was CdtB, but ferri-salmycin was not commercially available.

Assuming the 25% CdtB protein had unfolded hardly changed the calculated dissociation constants (see Appendix A.4) except for ferri-coelichelin (fourfold increase). If one assumed that only 50% of CdtB was folded, the model as determined by Equation 2 was unable to fit the data effectively. Despite the uncertainty of finding out how much active protein is in the solution, and taking the upper limit as the concentration determined by Bradford assay, the data does indicate unequivocally that the binding of CdtB to ferrioxamine B and ferri-coelichelin was much stronger than with the other ferri-siderophores tested.

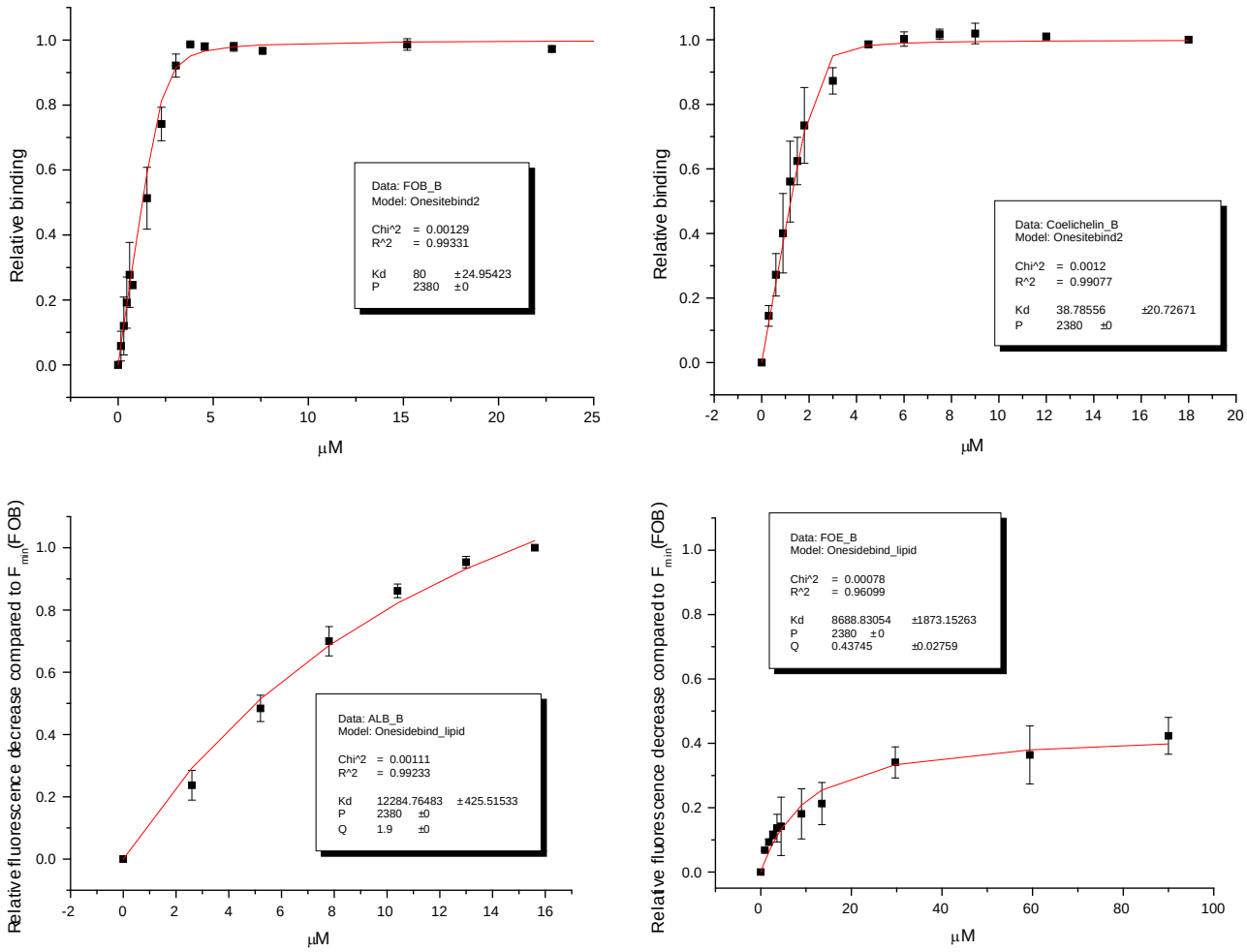


Figure 4-11: The relative binding of ferri-siderophores to a CdtB solution (78 µg/mL).

(Top left) ferrioxamine B; (Top right) ferri-coelichelin. The relative fluorescence decrease compared to F_{min} of ferrioxamine B bound to CdtB of (Bottom left) ferri-albomycin; (Bottom right) ferrioxamine E.

Table 4-5: Estimated dissociation constants (in µM) for ferri-siderophores from CdtB all protein was active in binding (100%) and only 75% was active in binding.

	CdtB (100 %)	CdtB (75 %)
Ferrioxamine B	0.080 ± 0.024	0.100 ± 0.056
Ferrioxamine E	8.7 ± 1.9	8.9 ± 1.8
Ferrichrome	77 ± 3	77 ± 3
Ferri-coelichelin	0.039 ± 0.021	0.172 ± 0.023
Ferri-albomycin	12.3 ± 0.4	12.6 ± 0.4

4.6 : Circular dichroism spectroscopy of DesE, CchF and CdtB proteins

Previous studies of conformational change upon binding of various ferri-siderophores of the *E. coli* periplasmic binding protein FhuD[28,89] and the *S. aureus* lipoprotein receptor FhuD2[38] found there was little effect. These studies involved X-ray crystallography and small-angle X-ray scattering in solution respectively. To investigate whether DesE, CchF and CdtB undergo conformational change on binding to ferri-siderophores, circular dichroism spectroscopy was used to investigate the secondary structure before and after binding. Circular dichroism (CD) measures the difference in absorbance between left and right circularly polarized light[83], and different secondary structural features such as α-helices and β-sheets can produce either positive or negative signals from 180 – 260 nm wavelengths of light due to their chiral nature.

The three proteins, DesE, CchF and CdtB were dissolved in 10 mM sodium phosphate buffer, and an excess of ferri-siderophore (either ferrioxamine B, ferrioxamine E or ferri-coelichelin) was added. CD spectra were measured at 21 °C from 190 nm to 260 nm, with the spectrum of the buffer subtracted from the spectra of the proteins. The results were processed in MS Excel with the baseline of buffer subtracted from the spectra.

These experiments show that the DesE (Figure 4-12) and CchF (Figure 4-13) proteins were fully folded with prominent α-helical content, but the spectrum of CdtB (Figure 4-14) seemed to be a combination of fully-folded and unfolded protein, with the latter being up to 25% in abundance[90] . This was indicated by a lack of a positive signal at 195 nm and the strong negative signal at 200 nm, which correspond to the CD spectrum of a random coil. The negative signals at 206 nm and 220 nm were indicative of α-helices in fully-folded proteins. These results, together with the early elution from the gel filtration column (earlier than the more massive proteins DesE and CchF - indicative of the larger hydrodynamic radius of the unfolded protein) and the gradual decrease of fluorescence upon standing at room temperature suggest that CdtB is unstable.

Addition of binding ferri-siderophores that bind tightly to the receptors did not change the CD spectra of the siderophore binding proteins DesE, CchF and CdtB significantly, indicating little change in secondary structure. This result was not unexpected considering the X-ray crystal structures of the *E. coli* ferric hydroxamate binding protein FhuD and *E. coli* vitamin B_{12} binding protein BtuF did not alter significantly on co-crystallization with their substrates. This was further confirmed from the X-ray crystallographic structures of apo-DesE and DesE-ferrioxamine B as discussed in the next chapter. The results of CdtB with its substrates ferrioxamine B and ferri-coelichelin ruled out the possibility that the reduction of fluorescence on addition of these ferri-siderophores was due to catalysis of protein unfolding; the CD spectrum hardly changed on addition of these ferri-siderophores, indicating little change in the secondary structure.

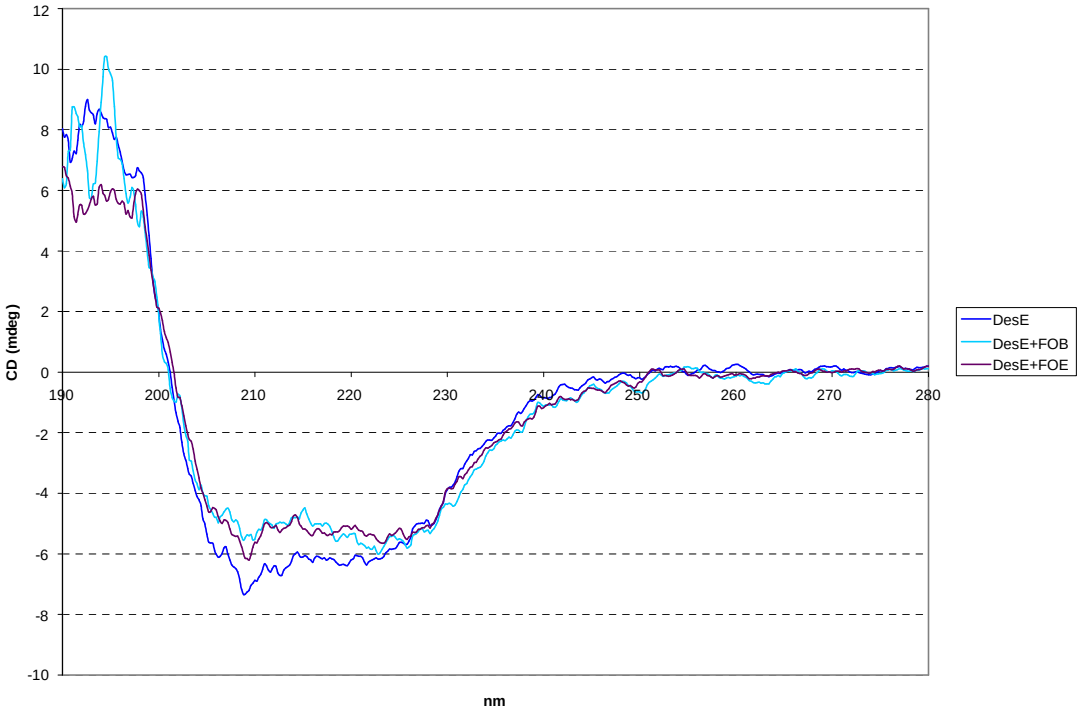


Figure 4-12: CD spectra of DesE with ferri-siderophores ferrioxamine B and ferrioxamine E.

80 µg/mL protein and 5 µM of ferri-siderophore were used in each case. Ferri-siderophores added: FOB (ferrioxamine B) and FOE (ferrioxamine E).

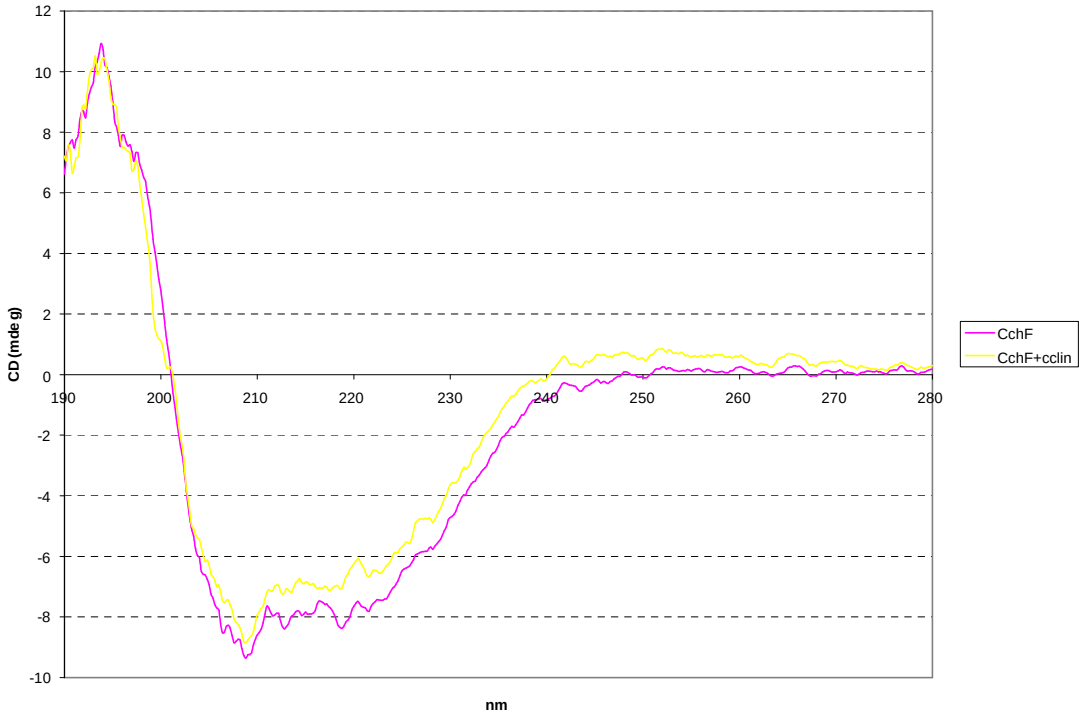


Figure 4-13: CD spectra of CchF with ferri-coelichelin.

80 µg/mL protein and 5 µM of ferri-siderophore was used. Ferri-siderophores added: cclin (ferri-coelichelin).

Figure 4-14: CD spectra of CdtB with ferri-siderophores ferrioxamine B and ferri-coelichelin.

80 µg/mL protein and 5 µM of ferri-siderophore were used in each case. Ferri-siderophores added: FOB (ferrioxamine B) and cclin (ferri-coelichelin). The absence of signal at 195 nm and the negative signal at 200 nm are indicative of the presence of random coils.

4.7 : Conclusions and future work

In the model of siderophore-mediated iron acquisition in *S. coelicolor* A3(2) proposed by Barona-Gómez *et al.*, the siderophores desferrioxamine B, desferrioxamine E and desferri-coelichelin are produced and are exported out of the cell. These bind to extracellular ferric iron to form their ferric complexes. The ferric complexes bind to the siderophore binding proteins DesE, CchF and CdtB, with DesE binding to both ferrioxamines B and E, CchF binding to ferri-coelichelin and CdtB binding to ferrioxamine B.

The fate of the DesE-ferrioxamine E complex seems to be straightforward, with the siderophore being transported via a yet unknown permease-ATPase complex into the cell, and it is assumed this is also the fate of the DesE-ferrioxamine B due to the similarity of the complexes formed. The fate of the CchF-ferri-coelichelin complex also seems simple enough, with the siderophore being transported via the permeases CchC and CchD and ATPase CchE also encoded in the coelichelin biosynthetic gene cluster. The fate of the CdtB-ferrioxamine B complex seems to be straightforward, with the permease-ATPase CdtAC importing the siderophore into the cell.

The xenosiderophores ferrichrome and ferri-albomycin were found to bind DesE. The fate of DesE-ferrichrome seems to be identical to that of DesE-ferrioxamine E, but the fate of the DesE-ferri-albomycin is unclear as *S. coelicolor* is not sensitive to this antibiotic. There may be several explanations for this: DesE-ferri-albomycin is transported into cells but do not kill them, or the DesE-ferri-albomycin binding may be too weak for its efficient uptake.

The most surprising aspect of the research was that CdtB also bound strongly to ferri-coelichelin, as the results of Barona-Gómez *et al.* suggest the CdtABC system is not essential for ferri-coelichelin uptake in *S. coelicolor*. As the CdtB protein was unstable in solution, this result should be treated with caution: for example, the

protein may be able to slightly unfold to cause a change in the binding site, which can now accommodate ferri-coelichelin, whilst the stable protein in *S. coelicolor* is unable to unfold to accommodate ferri-coelichelin. This could be investigated by comparing the kinetics of the formation of the CdtB-ferri-coelichelin to CdtB-ferrioxamine B complexes or other structural techniques.

If ferri-coelichelin does bind to fully-folded CdtB, does this mean the gene cluster *cdtABC* is involved in ferri-coelichelin transport? Barona-Gómez *et al.*[53] showed that non-siderophore producing mutants of *S. coelicolor* without the complete coelichelin (*cch*) gene cluster were unable to grow when supplemented with ferri-coelichelin (but grew on ferrioxamine B supplementation). The concentration of ferri-coelichelin added to the media was not determined in the study, hence it was not determined if the supplementation was at the same concentration as ferrioxamine B. But these concentrations of ferri-coelichelin stimulated growth on non-siderophore producing mutants of *S. coelicolor* containing most of the *cch* cluster (except for the NRPS gene *cchH*), hence the *cdtABC* cluster may not be essential for ferri-coelichelin uptake.

The genes that are essential to ferri-coelichelin uptake are very likely to be the genes that encode the CchCDEF ABC transporter system, but are all of the genes essential? It could be possible that CdtB-ferri-coelichelin could interact with the permeases CchC and CchD to transport ferri-coelichelin into *S. coelicolor*, which could be investigated by deletion of the *cchF* gene.

Another possible explanation of why the CdtABC system may be unable to import ferri-coelichelin is that the CdtB-ferri-coelichelin complex is not recognised by the CdtAC complex, whilst the CdtB-ferrioxamine B complex is recognised and the siderophore transported. The bound siderophore may have a large role in recognition of lipoprotein receptors by ABC transporters, which there is strong evidence in the literature.

Sebulsky *et al.*[38] reported mutants of the gene *fhuD2* gene (encoding a *Staphylococcus aureus* ferric-hydroxamate lipoprotein receptor FhuD2) which prevented the growth of *S. aureus* H431 on media supplemented with Fe(III)-

hydroxamates. The double mutation E97A/E231A was found to cause a profound decrease in growth of *S. aureus* H431 supplemented by ferrioxamine B and several other ferri-siderophores compared to *S. aureus* H431 with the wild-type *fhuD2*. However, this mutation had little effect on growth when supplemented by ferrichrome. The wild type and the E97A/E231A mutants of FhuD2 were overproduced in *E. coli* and were found to have almost exactly the same dissociation constants from ferrioxamine B, ferrichrome and the other ferri-siderophores. Thus, these mutations did not affect the siderophore binding site of FhuD2. It was concluded that ferrichrome could assist with the recognition of the receptor-ferri-siderophore complex by the permeases ABC transporter. Ferrichrome was proposed to compensate for the loss of the two glutamate residues that had a detrimental effect on ferrioxamine B uptake. Ferrichrome could do this in one of two ways: the siderophore interacts directly with the permease, or ferrichrome binding induces a different conformational change in FhuD2 than ferrioxamine B binding. The latter was investigated using small angle X-ray scattering, which showed little difference in the conformational change of FhuD2 on binding to ferrichrome compared to ferrioxamine B.

In support of the hypothesis that the siderophore interacts directly with the permease, an X-ray crystallographic structure[26] of the *E. coli* vitamin B_{12} periplasmic binding protein BtuF (analogous to the siderophore binding proteins) in complex with its ABC transporter permease-ATPase BtuCD, was proposed to show the $BtuC_2D_2F$ complex after translocation of vitamin B_{12}. The structure of BtuF without the vitamin, analogous to the siderophore binding proteins, was found with the substrate binding site directly interacting with the BtuC permease dimer, suggesting that the vitamin may play a role in the interaction of BtuF with the permease BtuC. Also two glutamate residues of BtuF were found also to interact with the BtuC permease dimer, which the study suggested also played a large role in recognition of the BtuF-vitamin B_{12} complex by BtuC as suggested by Sebulsky *et al.* for FhuD2.[38]

Another possible explanation of non-transport of ferri-coelichelin by the CdtABC system is that the CdtB protein is present in an “open” or “closed” form in solution. X-ray crystal structures of the *E. coli* SBP FitE[91] found an open and closed form of the protein, analogous to the *E. coli* open BtuF form when bound to the BtuCD

dimer[26] and apo-BtuF[27] respectively. Ferrioxamine B may bind to the closed form of CdtB, whilst ferri-coelichelin may bind to the open form, but only the closed, substrate-bound form of CdtB is recognised by CdtAC, hence only ferrioxamine B is taken up.

CdtB appears to discriminate between two very similar molecules: ferrioxamine B and ferrioxamine E; CdtB has more than 100 times the affinity for the former than the latter. Finding the molecular basis of this amazing specificity could be achieved by solving the structure of the protein by X-ray crystallography or by homology modelling of proteins of similar sequence.

Chapter 5: X-ray crystallographic structures of apo-DesE and DesE in complex with ferrioxamine B

5.1 : Introduction

The research in this section was carried out in collaboration with the group of Professor James Naismith of the University of St. Andrews. We provided the *E. coli* pPP004 plasmid which overexpresses the gene encoding the N-terminal truncated His_6-DesE. We also provided the ferrioxamine B used in the co-crystallization experiments which was prepared by reaction of the mesylate salt with an equimolar amount of $FeCl_3.7H_2O$ and subsequent freeze-drying. The purification, crystallization, X-ray diffraction and structure determination of apo-DesE and ferrioxamine B-DesE was carried out by researchers in St. Andrews.

5.2 : Crystallography details

Apo-DesE was crystallized by the hanging drop method using 0.1 M Tris-HCl (pH 7.0) with 3.2 M sodium malonate. The phasing data was obtained by soaking in the above buffer plus 0.5 M sodium bromide. The crystal for DesE-ferrioxamine B was obtained by soaking crystals of apo-DesE in a 1 mM solution of ferrioxamine B. The crystals were cryogenically frozen and the diffraction data was obtained at the European Synchrotron Radiation Facility (Grenoble, France). The resolution of the crystal structures of apo-DesE and DesE-ferrioxamine were 1.5 Å and 1.96 Å respectively.

5.3 : The structure of apo-DesE

The structure of *S. coelicolor* DesE is very similar to several bacterial siderophore and vitamin B_{12} binding proteins. In particular, DesE has a very similar structure to FeuA, a ferric-bacillibactin lipoprotein receptor protein from the Gram-positive bacterium *B. subtilis*[34,35]. DesE also shares many structural similarities with the vitamin B_{12}-binding protein BtuF from *E. coli* and the siderophore-binding periplasmic lipoprotein CeuE from the Gram-negative bacterium *Campylobacter*

jejuni[36], such as a rigid backbone of three α-helices bridging two β-sheet regions, whose loops form part of the substrate binding cleft (Figure 5-1).

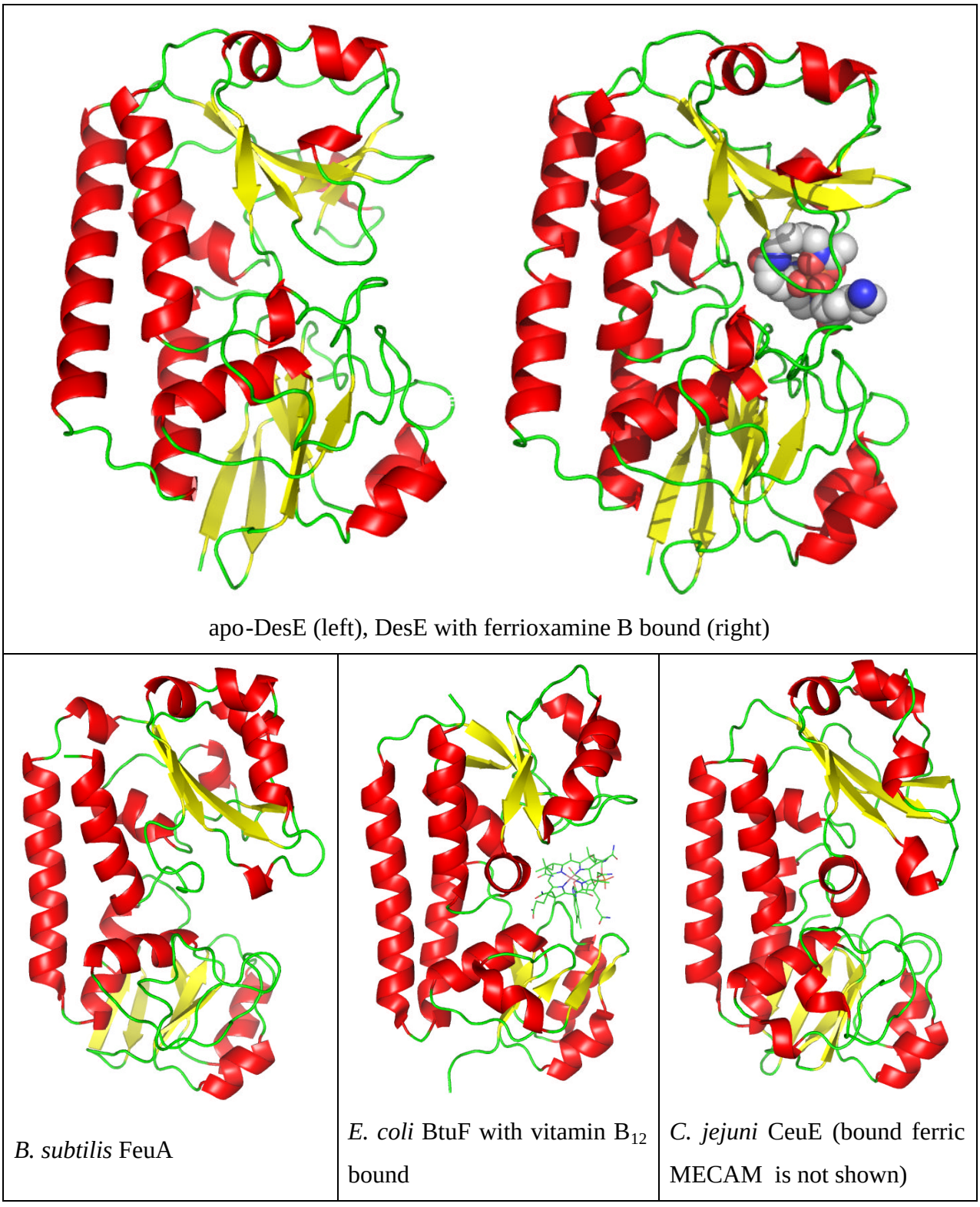


Figure 5-1: Comparison of the structure of apo-DesE and DesE with ferrioxamine B bound to other bacterial siderophore and vitamin B_{12} binding proteins.

5.4 : The structure of DesE with ferrioxamine B

The co-crystallization of the siderophore ferrioxamine B bound to DesE (Figure 5-1: top right) confirmed the affinity of the protein for the ferri-siderophore as indicated by the fluorescence quenching experiments described in Chapter 4. The co-complex structure identified tryptophan residues in the ferrioxamine B binding site which were close to or may directly interact with the ferri-siderophore (Figure 5-2), which were quenched very strongly in the fluorescence quenching experiments described in Chapter 4.

5.4.1 : Conformational change in DesE on binding of ferrioxamine B

DesE is related to the maltose binding protein (MBP) family of receptors involved in ABC importers[25]. MBP is a well-studied binding protein whose fold changes markedly on substrate binding. By contrast, the fold of siderophore-binding proteins has been observed to change very little on substrate binding. To examine whether DesE conforms to this paradigm, the structures of apo-DesE and DesE-ferrioxamine B were superimposed using the MOE package (Figure 5-3). This showed that the *apo*- and *holo*- forms of the protein have very similar structure (although part of the N-terminus of DesE was not observed in the crystal structures, presumably because it was disordered in the crystals).

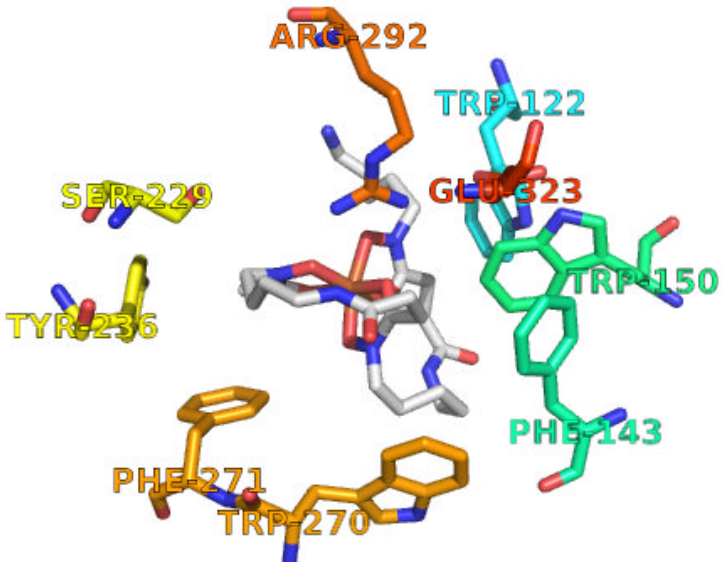


Figure 5-2: The ferrioxamine B binding site of DesE. The siderophore is in the centre.

5.4.2 : Residues involved in ferrioxamine B binding

As well as three tryptophan residues (W122, W150 and W270) that line the ferrioxamine B binding site, there was close contact of the ferric siderophore with the aromatic residues F143, F271 and Y236. The two residues S229 and E323 complete the binding pocket, but do not make close contact with ferrioxamine B. By far the most intimate contact between DesE and ferrioxamine B involves the R292 residue. Two of the hydrogen atoms of the guanidinium group in the side chain of this residue point towards to the hydroxamate oxime oxygen atoms of the ferri-siderophore, which may indicate the presence of hydrogen bonds. A close interaction between an arginine residue in *E. coli* FhuD to ferri-albomycin, ferric coprogen and ferric coprogen was also observed in X-ray crystal structures[28]. This interaction may be very important in stabilizing the complex, enabling the ligand to bind to the protein at nanomolar concentrations.

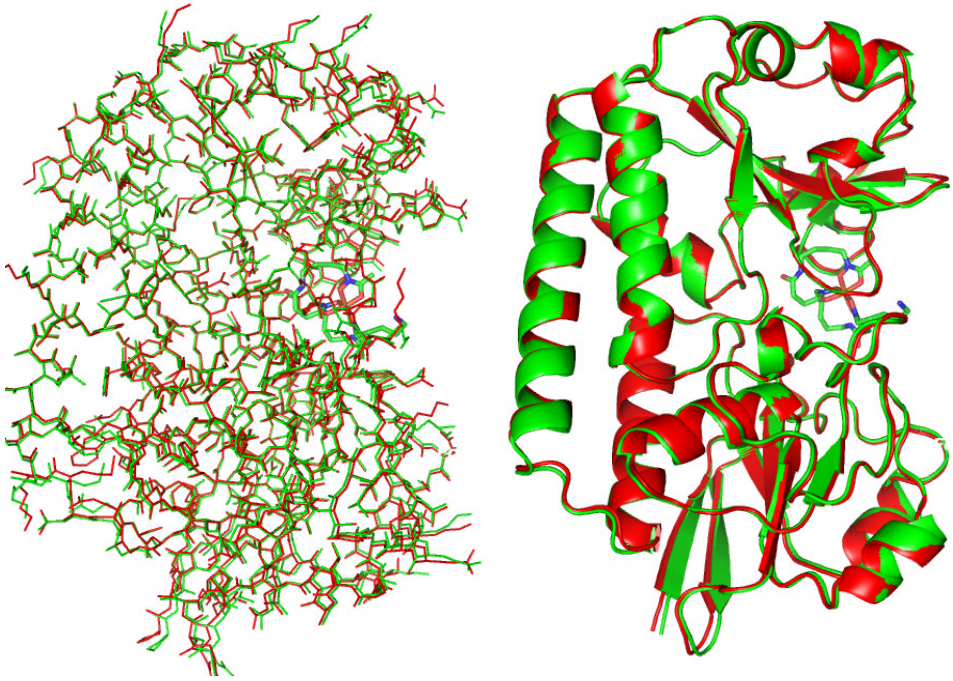


Figure 5-3: Superposition of DesE and DesE-ferrioxamine B.

Left: Stick models of apo-DesE (red) and DesE with ferrioxamine B bound (green). Right: Cartoon representations of apo-DesE and DesE-ferrioxamine B superimposed.

5.4.3 : The absolute configuration of bound ferrioxamine B

Free ferrioxamine B is a racemic mixtures of Λ and Δ isomers, which is confirmed by the observation that it crystallizes as both these isomers[72,92]. There are 16 possible geometrical and optical isomers of ferrioxamine B (which is also the case with the ferric complexes of linear hydroxamate siderophores) as the hydroxamate groups are inequivalent[11]. According to IUPAC rules, the first step in determining the configuration of ferrioxamine B is to determine whether the complex is of the left- or right handed propeller (Λ or Δ respectively). The next step is to give each hydroxamate group a number, starting from the group closest to the pendant NH_3^+ group in the case of the ferrioxamines. The numbers are assigned anti-clockwise and clockwise in Δ or Λ isomers respectively. Next, one must orient the molecule such that the hydroxamate groups read 1, 2, 3 when read anticlockwise in the case of Δ isomers (clockwise in the case of Λ isomers), then determine whether the oxime nitrogen is above (towards the page) or below (further into the page) than its corresponding carbonyl carbon in the first hydroxamate group. If the nitrogen is below the carbonyl carbon the complex is denoted N, otherwise it is denoted C. The hydroxamates 2 and 3 are compared with the first hydroxamate, in particular whether the orientation of the oxime nitrogen and carbonyl matches. If they do match, it is denoted *cis*, otherwise it is *trans*. The eight isomers of the Δ configuration are illustrated in Figure 5-4. The absolute configuration of ferrioxamine B in the binding pocket of DesE was sufficiently determined from the electron density map (Figure 5-5), and is of an unusual configuration (Δ-N-trans-cis: Figure 5-6). Ferrioxamine B bound to *E. coli* FhuD adopted a Λ-C-*trans*-*cis* configuration (although the reference says it is Λ-*cis*) [28].

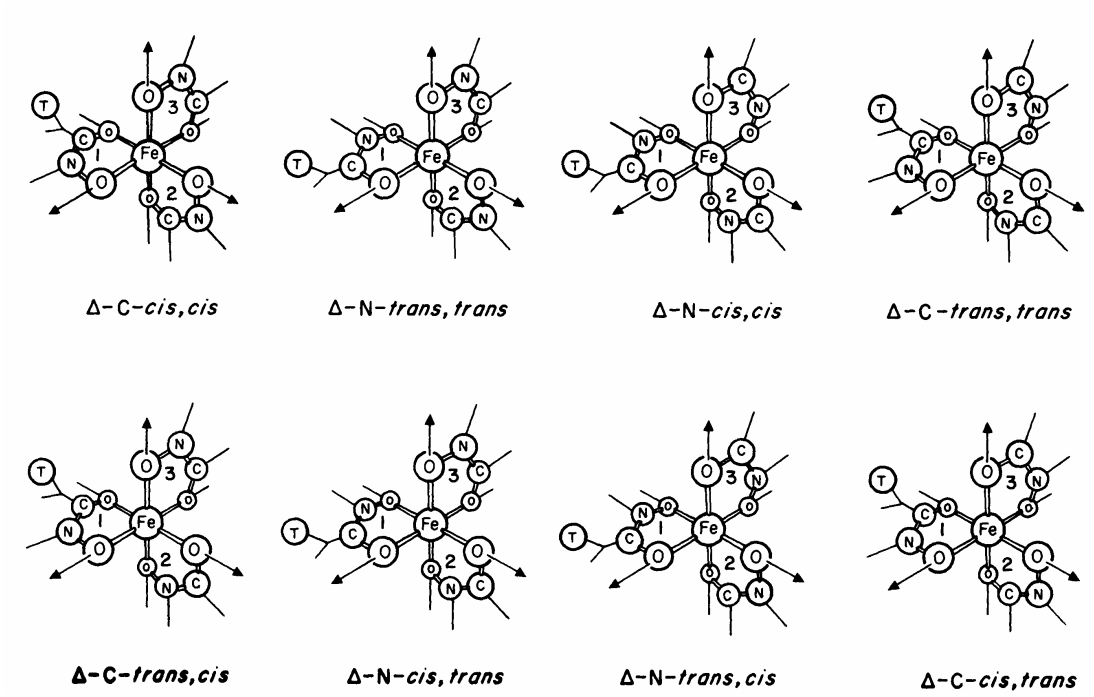


Figure 5-4: The eight geometrical isomers of a ferric-*tris*-hydroxamate with inequivalent hydroxamates in a Δ configuration.

In the case of ferrioxamine B, hydroxamate 1 corresponds to the group closest to the free amino group. The equivalent Λ isomers are the mirror images of the corresponding Δ isomer. Figure from [11].

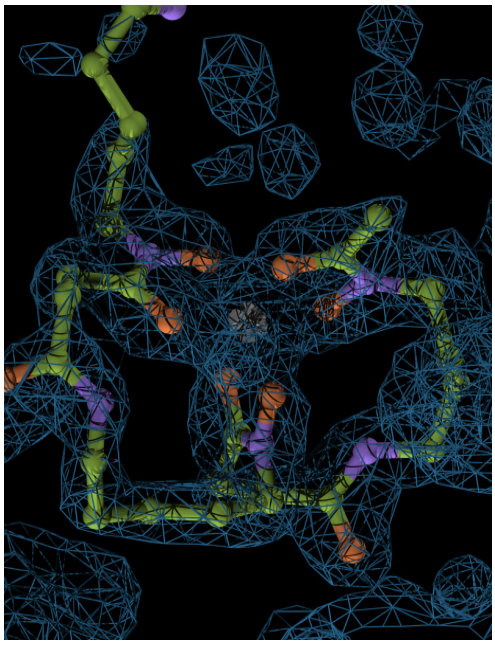


Figure 5-5: The electron density map of the DesE-ferrioxamine B around the ferri-siderophore.

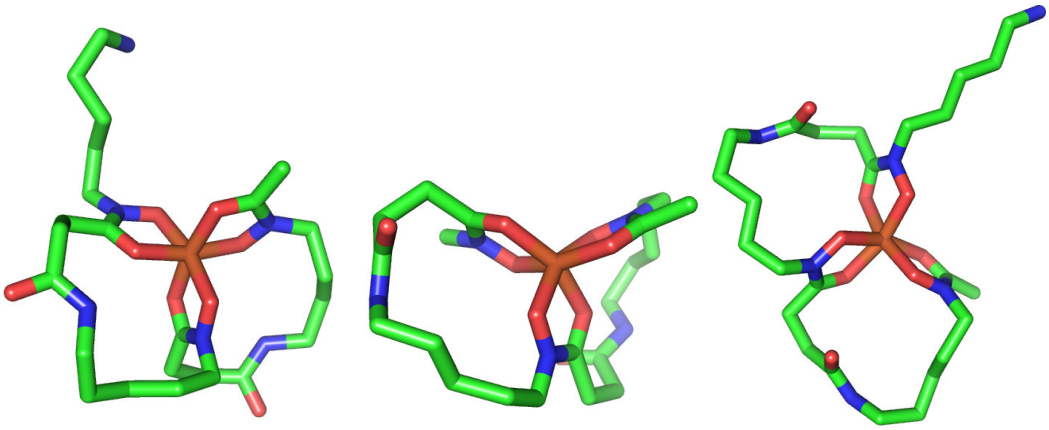


Figure 5-6: Ferrioxamine B isomers.

From left to right: Ferrioxamine B bound to DesE (configuration Δ-N-*trans-cis*); Ferrioxamine B bound to FhuD (configuration Λ-C-*trans-cis*); free ferrioxamine B crystal structure (configuration Δ-N-*cis-cis*)[92].

Chapter 6: Molecular modelling of ferric hydroxamates

6.1 : Introduction

The aim of this chapter was to construct a molecular mechanics force field model that accurately reproduced bond lengths, angles and the energetics of trigonal twists of ferric-*tris*-hydroxamate complexes such as siderophores. This model may ultimately be used for molecular dynamics simulations of siderophore binding proteins.

6.2 : Hydroxamate complexes

To understand more about the nature of siderophore complexes it is important to consider the interactions between the siderophore ligands and the Fe(III) ion. This ion has a high-spin d^5 electronic configuration in hydroxamate complexes – a half-filled d-electron shell – and consequently there are no ligand field effects caused by an uneven distribution of electron density in this shell, and form thermodynamically stable hexacoordinate octahedral complexes, mainly with oxygen donors.

Siderophores, such as the ferrioxamines and ferrichromes, chelate Fe(III) to fom *tris*-hydroxamate complexes. The lone pairs of the deprotonated oxime and the carbonyl oxygens of the hydroxamate form coordinate bonds with Fe(III) to form a complex which is structurally and thermodynamically very stable[11], due to the delocalization of the deprotonated hydroxamate on Fe(III) binding[93], whose resonance forms are shown in Figure 6-1.

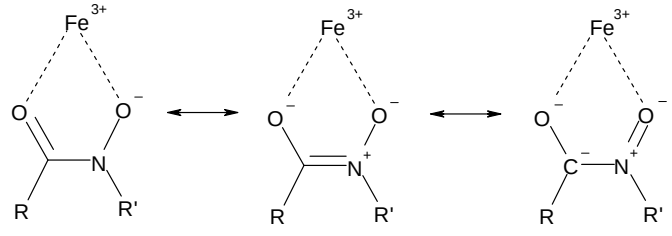


Figure 6-1: The resonance forms of the hydroxamate-Fe^{3+} five membered ring[93].

The formation constants of hydroxamate siderophores with Fe(III) vary (Table 6-1), but in general their affinities are less than catecholate-type siderophores such as enterobactin[11]. There is a variety of influences on this property, including the outer structure of the siderophore or the entropic change on binding.

Table 6-1: Formation constants[11], β_{110}, of ferric tris-hydroxamate siderophores for the reaction $Fe^{3+} + L^{n-} \rightarrow FeL^{3-n}$**, where** $\beta_{110} = \frac{[FeL^{3-n}]}{[Fe^{3+}][L^{n-}]}$.

Siderophore	$log_{10}\ \beta_{110}$
Ferrioxamine E	32.5
Coprogen	30.2
Ferrioxamine B	30.5
Ferricrocin	30.4
Ferrichrysin	30.3
Ferrichrome A	32.0
Ferrichrome	29.1

The shape of the ferric complex is very important in the recognition of microbes in determining the uptake of these complexes, and in particular the local coordination environment of Fe(III): for example, the earlier chapters demonstrated that three siderophore binding proteins from *Streptomyces coelicolor* A3(2) had different specificities to the ferric-*tris*-hydroxamates ferrioxamine E, ferrioxamine B, ferri-coelichelin, ferrichrome and albomycin. As well as the overall shape of the complex being important for recognition, the absolute configuration of the hydroxamate ligands around the Fe(III) ion may also play a role in recognition[11].

6.3 : Survey of X-ray Crystallographic Structures

To construct a molecular model which was effective at modelling ferric *tris*-hydroxamates, the X-ray crystallographic structures of the complexes were used as the minimal energy conformers of the *tris*-hydroxamates.

The crystallisation of siderophores with Fe(III) is the orthodox method for determining the structure of complexes. As such complexes are small (about 100

atoms), the resulting crystal structures were usually very accurately determined without much disorder, and, in principle, should correspond to a energy minimized configuration, but may also co-crystallise with another conformer if they were very similar in energy, as in the case of ferric tris-acetohydroxamate, where both the *fac* and *mer* isomers co-crystallized[71] .

The weakness of this information is that only one conformer is usually found in the structurally static crystal. If there were different conformers in solution with similar energies only one would be found – there is no information about any flexibility of the molecule. Also the conditions of the crystallisation are usually different to the physiological environments they are naturally found in.

A promising approach to obtain a solution-state structure is to use nuclear magnetic resonance (NMR) on a metal-siderophore complex where Fe(III) is replaced by a non-paramagnetic ion of similar atomic radius such as Ga(III) or Al(III). The structure can be obtained for small conformers, but if there was more than one conformer the structure obtained would be an average of all possible conformers. This would introduce disorder into the system, but would give a more realistic picture of the flexible nature of their structures.

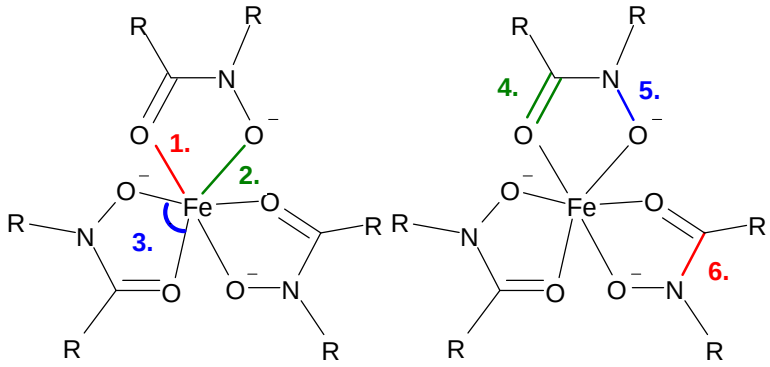
The Cambridge Crystallographic Structural Database (CCSD) was used to obtain ferric tris-hydroxamate crystal structures. There were 27 ferric tris-hydroxamate complexes present in the Database, including binuclear complexes such as ferric alcaligin[94].

All the complexes in the Database were close to octahedral in the coordination environment of the Fe(III) ion, with most exhibiting the *fac* geometry, where the three carbonyl oxygen atoms were on a plane opposite the three oxime oxygen atoms. There was a significant difference of Fe – O bond lengths of 0.055Å on average between the formally negatively charged oxime oxygen O(N) to that of the lone pair bond of the carbonyl oxygen, O(C), with the Fe – O(N) being shorter than Fe – O(C) (Table 6-2). This was to be expected from the electrostatic attraction between the O- and formally positively charged Fe^{3+} ion, but in ferrioxamine E, this

difference was 0.1Å, which could be due to the influence of the cyclic structure of the siderophore.

Table 6-2: Average bond lengths and angles from 27 ferric tris-hydroxamate metal centres from X-ray crystallographic structures.

		Average	Standard Deviation
1.	Fe – O(C) bond length	2.039 Å	0.022 Å
2.	Fe – O(N) bond length	1.980 Å	0.017 Å
3.	O(N) – Fe – O(C) bite angle	78.5°	0.66°
4.	C = O bond length	1.278 Å	0.0125 Å
5.	N – O bond length	1.377 Å	0.0135 Å
6.	C – N bond length	1.319 Å	0.0256 Å
	Trigonal Twist	41.7°	3.7°



The O(N) – Fe – O(C) bond angle within one hydroxamate (the ligand bite angle) is much smaller than 90°, hence hydroxamate complexes were not rigorously octahedral with respect to the oxygens surrounding the Fe(III) centre, but closer to D_{3d} or C_3 symmetry. The metal-hydroxamate ring is planar, probably due to the influence of the delocalization of the π-system around the hydroxamate.

6.3.1 : Trigonal distortion: measuring trigonal twist

In all the tris-hydroxamate structures, there was little variation in bond lengths and bite angles around the metal centre. The main distortion from the unrestrained ferric tris-hydroxamate complexes caused by the tether of the siderophore structures is the

departure from a roughly octahedral complex (approximately D_{3d}) to a trigonal prismatic C_{3v} structure (see Table 6-2). Most hydroxamate complexes are in between these two extremes.

This distortion is known as the trigonal twist, and is measured as follows. Viewing from a pseudo-C_3 axis of a tris bidentate complex, the ligands' donor atoms in a perfectly octahedral complex would look like two triangles whose vertices line up to the midpoints of the other triangle's edges in a staggered formation, whilst in a trigonal prism, a configuration with no trigonal twist, these triangles would be eclipsed (Figure 6-2). This can be quantified by measuring the angle α (the twist angle) which is 0° for a trigonal prism and 60° for a perfectly octahedral complex[93,95].

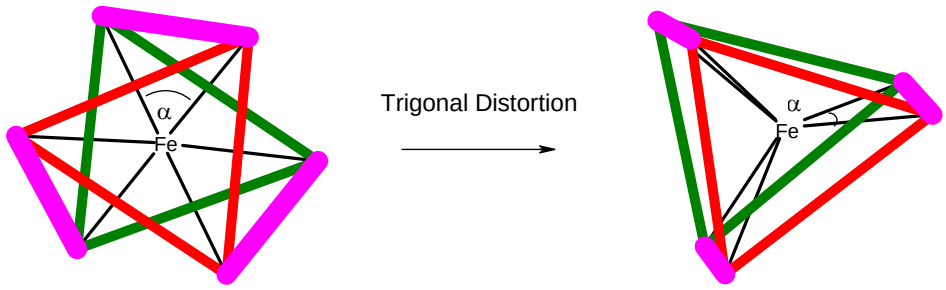


Figure 6-2: Defining the twist angle of a bidentate ligand.

The red and green lines represent the triangles seen from above with the C_{3V} axis pointing out of the page. The thick pink lines represent one bidentate ligand (e.g. one hydroxamate). In trigonal distortion, α is reduced from an "octahedral" complex (left) to a trigonal prismatic complex (right).

To measure the trigonal distortion in *fac* ferric hydroxamate complexes, the vertices of one of the triangles were formed by the carbonyl oxygen atoms, and the other triangle had the oxime oxygen atoms of the hydroxamate (for meridional isomers any two carbonyl and one oxime oxygen would form a suitable triangle if they are from different ligands).

The pseudo-C_3 axis is constructed by the vector that passes through the centroids of both triangles. Then the twist angle α for a *fac* isomer is simply the angle between the vector of a carbonyl oxygen and its centroid, and the corresponding oxime

oxygen and centroid of the same ligand, calculated by application of the dot product. The value calculated by this method may be inaccurate if the planes containing the carbonyl oxygen atoms and the oxime oxygen atoms were not close to parallel, or the metal atom is significantly displaced from the C_3 axis.

In reality, the trigonal twist of the complexes was rarely as large as 60°, and the optimal trigonal twist is largely dependent on the ligand bite: the O···O distance in a hydroxamate group divided by the M···O bond length. The trigonal twist of ferric hydroxamates can be estimated from the ligand bite[96], and a relation between twist angle and ligand bite has been presented:

$$\text{Twist angle} = -73.9 + 94.1\ b$$

where *b* denotes the ligand bite. There was a markedly greater variation in trigonal twists between complexes than bond lengths and angles, and seemed to be the main distortion caused by the constraints of a backbone structure of a siderophore compared to an unconstrained tris-hydroxamate. For ferric *fac* tris-acetohydroxamate[71] the trigonal twist was 44.6°, whilst for ferrioxamine B and ferrichrome they were 44.1° and 42.9° respectively.

The *mer* isomers in the database, in particular *mer* tris-acetohydroxamate, neocoprogen I and ferric alcaligin had lower trigonal twists than the *fac* isomers, with an average of 37.0°. Also, some ferric-*tris*-hydroxamate complexes with bulky phenyl groups showed low trigonal twists of 35°. This could be due to the electronic effects due to the aromatic groups or possibly due to crystal packing effects.

6.4 : Computational studies on transition metal complexes

Quantum mechanical (QM) methods are by far the most accurate method of determining structures and properties of small molecules, especially transition metal complexes. QM treats a molecule as a system of electrons and fixed, heavy, positively-charged nuclei. A plethora of information can be obtained from QM with high accuracy, such as absolute/relative energies, spectra and structures. The downside is the computational cost – a 100 atom system could take up to 2 weeks on

a 3.0 GHz processor – hence approximations have to be made, compromising accuracy.

In contrast, molecular mechanics (MM) methods represent a molecule as point charges with bonds modelled by springs between them in the simplest case. Note MM does not explicitly take into account the electronic properties of the complex, hence is an extreme approximation to a molecule. It is a highly empirical method, where properties of such as bond lengths, angles, torsional terms and partial charges for electrostatics have to be estimated to produce reasonable answers in a particular context. A particular model or parameterization is known as a force field. Despite these shortcomings, MM has been a very successful method for modelling macromolecules such as proteins and lipids, and results are comparable with QM for molecules containing main group elements.

6.4.1 : Density Functional Theory (DFT)

Quantum mechanical calculations are commonly used to calculate accurate properties of a molecule. In principle this is done by solving the time-independent Schrödinger wave equation for the system, where n is the number of electrons and nuclei:

$$H(\tau_1,\tau_2 \ldots \tau_n)\Psi(\tau_1,\tau_2 \ldots \tau_n) = E\Psi(\tau_1,\tau_2 \ldots \tau_n)$$

The Hamiltonian *H* describes the energetic interactions between the particles (nuclei and electrons) in the system whilst the waveform describes their spacial distrubution. The Hamilitonian is the sum of the kinetic energy terms and potential energy terms, *T* and *V*, where subscripts *A* and *i* are sums over nuclei and electrons respectively. M_A and Z_A are the relative mass and charge of nucleus *A* relative to an electron and ∇^2 is the Laplacian operator in three dimensions:

$$H = T + V = \underbrace{\left(-\sum_A \frac{1}{2M_A}\nabla_A^2 - \sum_A \frac{1}{2}\nabla_i^2 \right)}_{T} + \underbrace{\left(\sum_A \sum_i \frac{Z_A}{r_{Ai}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{r_{AB}} \right)}_{V}$$

This cannot be solved exactly for systems with more than one electron, hence to investigate larger systems, an approximation has to be made. The Born-Oppenheimer approximation assumes that the larger nuclei move much more slowly than electrons and so can be treated as static. Hence the Hamiltonian *H* becomes:

$$H = \underbrace{\left(\sum_A \frac{1}{2} \nabla_i^2 \right)}_{T} + \underbrace{\left(\sum_A \sum_i \frac{Z_A}{r_{Ai}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{r_{AB}} \right)}_{V}$$

where the kinetic energy term involving the nuclei disappears (because the nuclei are not moving) and the final term in the potential energy is constant. The wavefunction $\Psi(r_i; R_A)$ is now a function of only electron positions r_i with parameters R_A – the coordinates of fixed nuclei.

As this is an approximation, energies and spectra gained from this method are no longer exact, but the size of the error significantly reduces as the molecule becomes larger. To evaluate the minimal energy geometry, different nuclear positions are calculated, which give rise to different wavefunctions Ψ which give different energies *E*, until there exists a wavefunction which any variation of nuclear positions increases the energy.

One of the first approaches of QM is the Hartree-Fock method (HF)[97], where the complete wavefunction is approximated by a combination of one-electron wavefunctions. Each electron's potential is calculated by considering all the other electrons and nuclei, but it neglects electron correlation - it assumes that the movement of a particular electron is independent to the rest of the electrons (which is untrue as two electrons cannot occupy the same quantum state). Also, it is increasingly computationally expensive as the system becomes larger.

The electron exchange energy calculated by HF methods becomes increasingly inaccurate for heavier nuclei such as those in transition metals – instead the Møller-Presset peturbation[97] (MP2) method can be used, which reduces this error significantly. The only problem is the computational effort for MP2 is now five orders of magnitude instead of four.

The other popular approach is density functional theory (DFT)[97]. Instead of solving the many-body electronic wavefunction, the electron density is considered instead, which is a function of three variables instead of 3N variables (where N is the number of electrons) in HF theory. The lynchpin to this method is the Hohenberg-Kohn theorem, which states that the ground state electron density minimizes the total electronic energy of the system. The problem with this method is that it treats the system with non-interacting electrons hence electron correlation and electron exchange terms have to be added explicitly.

The performance of DFT methods has improved greatly during the 1990s where it can now compete with HF-based methods for accuracy. Geometries calculated using the local density approximation (LDA), where the initial estimate of the electron density around the orbitals is that it is completely homogeneous (a uniform electron gas), underestimates bond lengths, but these errors are smaller than the overestimation of bond lengths given by HF theory. These methods are combined in the Becke's hybrid method B3LYP[97], which is commonly used in many QM calculations on main group elements. Instead, using DFT methods with LDA with gradient corrected functionals gives similar results to MP2 but much faster, at three orders of magnitude, hence it is the most popular method for calculations on transition metal complexes.

6.4.2 : Molecular Mechanics

To deal with much larger complexes within a reasonable timescale, empirical methods such as molecular mechanics (MM) must be used, which is an effective method for very large molecules such as proteins. The strain energy *E* of a system of molecules in MM is calculated by:

$$E = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

where E_{str}, E_{bend}, E_{tors}, E_{vdw}, E_{el} and E_{cross} are the stretch energy between two atoms, angle bend strain between three atoms, torsional strain between four atoms, van der Waals strain between atoms, electrostatic interactions between atoms, and cross

terms (e.g. stretch bends) respectively. The stretch energy between two atoms A and B is given by the general formula:

$$E(R^{AB}) = k_2^{AB}(R^{AB} - R_0^{AB})^2 + k_3^{AB}(R^{AB} - R_0^{AB})^3 + k_4^{AB}(R^{AB} - R_0^{AB})^4 + \cdots$$

where the ideal bond length R_0^{AB} terms and force constants k_2, k_3 and k_4 have to be determined, and there are similar expressions for the other terms. A scheme of these constants is termed a forcefield. In many forcefields, for example Halgren's MMFF94 forcefield[98], the parameterisation of the forcefield is based on QM calculations on main group molecules. To find the minimum energy structure, the strain energy E is minimised.

The central idea of molecular mechanics is that these constants are transferable to other molecules. Most atoms in organic molecules adopt very similar bond lengths and angles depending on the geometry of the atom. Most C – H bond lengths are 1.06 to 1.10 Å in just about any molecule, with stretching frequencies between 2900 and 3300 cm^{-1}. This strategy is refined using different "atom types"- for example a sp^3 hybridised carbon atom in ethane would have different properties than a sp^2 hybridised carbon atom in ethene which would be reflected in the forcefield properties. In the MMFF94 forcefield[98], the ethane carbon would be denoted by the symbol C whilst C= would denote the ethene carbon. Other terms such as out-of-plane bends may be added to account for other effects (e.g. a nitrogen atom in an amide bond is planar contrary to many nitrogen-hydrogen compounds which are pyramidal) to refine the forcefield even further.

The main problem with this method is that it takes no account of the electronic environment of the atoms, which can cause problems if this is not taken into account beforehand by the forcefield applied.

6.4.3 : Molecular mechanics on Fe(III)

Transition metal complexes are usually hard to model in an MM scheme because of high coordination numbers, variable oxidation states, different spin states, and other

electronic effects such as Jahn-Teller distortions which are absent in molecules containing only light elements. To capture the behaviour of transition metals into an MM scheme, a number of models have been proposed, such as DOMMIMOE[73], SHAPES[99], which employs a Fourier transform description of the angular potential, VALBOND[100], which uses valence bond theory on transition metals, and MOMEC[101], a purely ligand-ligand repulsion approach.

The program used in this report, DOMMIMOE[73], offered capabilities which make it especially useful for handling transition metal complexes. The d-electrons around a transition metal can exert a strong influence on the energy and geometry in a complex. These are known as ligand field effects and can be incorporated into the MM scheme by adding a term called the ligand field stabilization energy (LFSE)[102].

To calculate the LFSE, a bond-centred form of ligand field theory known as the angular overlap model (AOM) is used, where the interaction energy of the ligand bonding and antibonding orbitals is calculated with respect to the probable d-electron density of the metal. The metal-ligand bond stretch potential function is modelled by a Morse function, which has a very similar shape to the realistic Lennard-Jones potential at short distances.

The positioning of the ligands around the metal centre is modelled by a purely repulsive interaction. If there is no ligand field effect, then this becomes a purely points-on-a-sphere model where the ligands are ideally arranged as far away as possible, which accounts for the mainly trigonal bipyramidal, octahedral and square antiprismatic geometries for most 5, 6 and 8 coordinate metal complexes.

Hydroxamate and other ferric siderophore complexes are almost entirely high-spin d^5 Fe(III). There is no ligand field effect hence the LFSE is zero. This meant the model can be parameterised by bond stretch terms between the Fe and hydroxamate O atoms, ligand-ligand repulsion terms between the hydroxamate O atoms, and an atomic partial charge scheme for electrostatic interactions. If Fe(II) binding has to be considered, for example in studies investigating the reduction of the Fe(III) centre to Fe(II) as in proposed models of iron release from siderophores once in the cell, explicit ligand field effects have to be considered for the ferrous complexes.

The other advantage of DOMMIMOE is that it is implemented into the popular molecular mechanics package Molecular Operating Environment[74] (MOE), which has a graphical user interface which facilitates construction and manipulation of molecules and also has an expandable pharmacological functionality with its own programming language called SVL (scientific vector language), where programs can be written to automate processes. Recently, molecular dynamics has been implemented into DOMMIMOE, hence ferric siderophores could be modelled with siderophore-binding proteins in an aqueous environment to obtain binding information.

6.4.4 : Previous computational studies on ferric hydroxamates

Considering the importance of siderophores in medicinal chemistry, there are surprisingly few computational studies on these small molecules. This is due to the complexes being too large for effective quantum mechanical treatment (>100 atoms) whilst being too small coupled with inadequate methods of treating the metal centre with MM (hence giving bad relative comparisons to known X-ray crystal structures).

Zinelabidine *et al.* (1993)[103] use an MM technique to model the Fe(III) ion in catecholate, hydroxamate and EDTA derivatives to calculate a rudimentary strain energy change when complexed with Fe(III). A model of ferric tris-acetohydroxamic acid and ferrioxamine E was constructed, and this was compared with the X-ray crystallographic structure of the latter[104].

Herrera et al (2006).[105] constructed an empirical forcefield in the program FORCES-2000 to model ferric-*tris*-hydroxamates. They parameterized the forcefield using both X-ray crystallographic data and *ab initio* calculations (Hartree-Fock) to model geometries of isomers of the ferric complexes of acetohydroxamic and N-acetohydroxamic acids, and calculate the relative energies of Δ, Λ, *fac* and *mer* isomers. They concluded the facial isomers were energetically favoured, but the energies of the meridional isomers were quite close, leading to the conclusion they could interconvert, especially in siderophores such as ferrioxamine E and ferrichrome although these siderophores were not tested in the model.

In Edwards and Myeni (2005)[106], quantum mechanical calculations (DFT and B3LYP) were used to justify XAFS experiments on ferric tris-(cis-acetohydroxamic acid) and ferrioxamine B to determine the Fe – O bond lengths in solution. In the study, they determine the protonation state of Fe tris-(cis-acetohydroxamic acid) as neutral except in very alkaline conditions, where the oxime hydrogen is deprotonated, and that ferrioxamine B has a formal charge of +1 in aqueous solutions.

Felder and Shanzer (2003)[107] have also developed an empirical force field for the modelling of ferri-siderophores such as ferrioxamine B and ferrichrome, but have not disclosed geometric and energetic information accompanying their results.

6.5 : Parameterization of Fe tris-hydroxamates in DOMMIMOE

6.5.1 : Aim of the parameterization

The first aim of the model was to produce a forcefield that accurately reproduces the bond lengths and angles of ferric *tris*-hydroxamate complexes, and also to reproduce the trigonal twist of the complexes, as it was the main distortion amongst the X-ray crystallographic structures (as summarized in Table 6-2) and also may be important for recognition by outer membrane receptors. It would be achieved by the search of suitable parameters for DOMMIMOE, in particular bond stretch terms between the Fe and hydroxamate O atoms, ligand-ligand repulsion terms between the hydroxamate O atoms and the charge scheme for electrostatics.

In the MMFF94 forcefield, the carbonyl oxygen in the hydroxamate is denoted by the symbol O=, and the oxime oxygen by OM. These symbols will be used from this point onwards in this chapter.

6.5.2 : DFT calculations on simple ferric-tris-hydroxamates

MM has the useful feature that a parameterization of a small complex is transferable to larger complexes, so if the parameters produce the desired behaviour for small

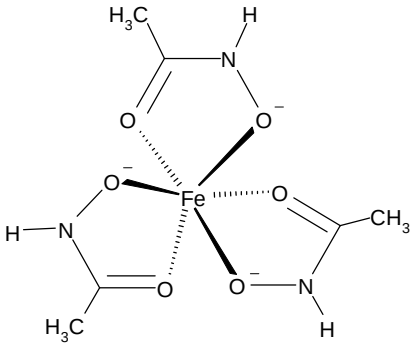
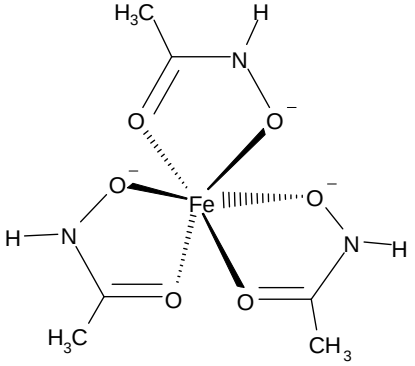
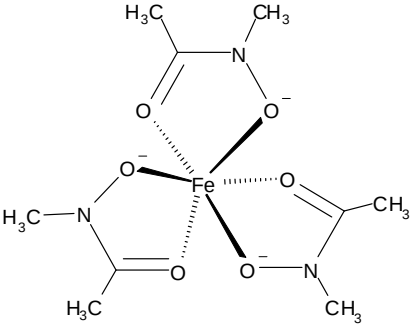
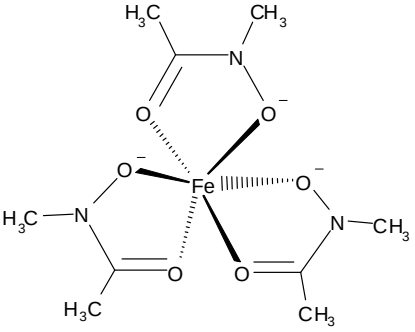
complexes they would give reasonable results for larger complexes. Hence, one must start on the simplest of ferric *tris*-hydroxamates, namely ferric tris-(cis-acetohydroxamic acid) denoted $Fe(hydr)_3$, and ferric tris-(cis-N-methyl-acetohydroxamic acid) denoted $Fe(Nmhydr)_3$, which were representative of atom types found in ferri-siderophores. To obtain accurate energetic information of these complexes, it was necessary to perform DFT (quantum mechanical) calculations to obtain another handle to optimize the molecular mechanics.

6.5.2.1 : **Facial and meridional isomer calculations**

Quantum mechanics calculations were performed on the following complexes: *fac* $Fe(hydr)_3$, *mer* $Fe(hydr)_3$, *fac* $Fe(Nmhydr)_3$ and *mer* $Fe(Nmhydr)_3$. There were X-ray crystallographic structures for the two isomers of $Fe(hydr)_3$, whilst $Fe(Nmhydr)_3$ was a theoretical complex. Quantum mechanics was used to optimize the structures of these four complexes, which in the case of the two isomers of $Fe(hydr)_3$ gave structures which were almost identical to the crystal structures. As the same method was applied to the $Fe(Nmhydr)_3$ complex, it was assumed this was an accurate reproduction of the structure.

The *fac* and *mer* isomers of $Fe(hydr)_3$ had an absolute energy difference of 1.31 kcal/mol, with the *fac* isomer being lower in energy (Table 6-3). This difference was very small, and goes some way in explaining why both isomers co-crystallized. By contrast, the *fac* isomer of $Fe(Nmhydr)_3$ was lower in energy by 4.01 kcal/mol than the *mer* isomer, which may explain why *fac* isomers are more commonly seen in metal-siderophore complexes than *mer* isomers.

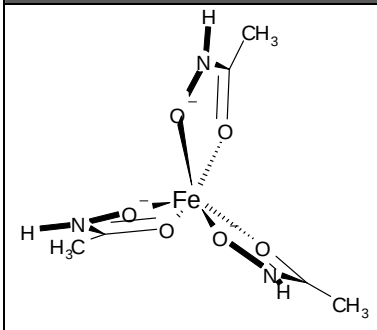
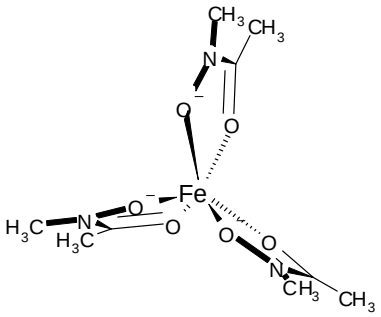
Table 6-3: The DFT energies of minimized structures of *fac* and *mer* isomers of simple ferric-*tris*-hydroxamates.

Structure	Description	DFT energy (kcal mol^{-1})
	*fac* ferric tris-(cis-acetohydroxamic acid)	-3960.32 (0.0)
	*mer* ferric tris-(cis-acetohydroxamic acid)	-3959.01 (+1.3)
	*fac* ferric tris-(cis-N-methyl-acetohydroxamic acid)	-5085.46 (0.0)
	*mer* ferric tris-(cis-N-methyl-acetohydroxamic acid)	-5081.45 (+4.0)

6.5.2.2 : **Facial isomers with 0° trigonal twist**

To estimate the energies required for a trigonal twist of these simple complexes, quantum mechanical calculations were performed on *fac* isomers of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ such that they had a trigonal twist value of 0°. This was achieved by constructing the complexes such that the planes of the hydroxamate groups were perpendicular to the planes formed by the oxime and carbonyl oxygen atoms, and then imposing C_{3v} symmetry when optimizing the structure of the complex (Table 6-4). The difference in energies between the minimized ("octahedral") *fac* isomers of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ and their C_{3v} minimized structures were 6.29 kcal mol^{-1} and 5.83 kcal mol^{-1} respectively.

Table 6-4: The DFT energies of minimized simple ferric-*tris*-hydroxamates with C_{3v} symmetry imposed.

Structure	Description	DFT energy (kcal mol^{-1})
	*fac* ferric tris-(cis-acetohydroxamic acid) C_{3v} minimized	-3954.03 (+6.3)
	*fac* ferric tris-(cis-N-methyl-acetohydroxamic acid) C_{3v} minimized	-5079.63 (+5.8)

6.5.3 : Parameterization of DOMMIMOE

The search for parameters in DOMMIMOE focussed on replicating the DFT structures of two simple ferric tris-hydroxamates: *fac* $Fe(hydr)_3$ and *fac*

$Fe(Nmhydr)_3$. The other focus was to reproduce the energy difference of minimized $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ with their C_{3v} symmetry minimized structures such that DOMMIMOE could be able to model trigonal distortion these ferric hydroxamates.

6.5.3.1 : **Morse function parameters**

The Fe – OM (Fe to oxime oxygen bond) and Fe – O= (Fe to carbonyl oxygen bond) bond stretch potentials in DOMMIMOE were modelled by the Morse function, which is given by the formula $E(Morse) = D[1 - e^{-\alpha(r-r_0)}]^2$ where r_0 denotes the reference bond length, D the depth of the potential 'well' and α the curvature of the well (Figure 6-3). As the Fe – oxygen distance approaches infinity, the Morse function tends to D, whilst the minimum is at r_0 with $E(Morse) = 0$. As the metal-ligand distance becomes shorter than r_0, the energy rises steeply, mimicking the effect of strong repulsion caused when atoms come too close. The curvature parameter α was important when considering bond vibrations and energies away from the minimal geometry. In the optimization of these parameters, the parameters α and D were set to 0.6 and 60 respectively (these parameters have little effect on the geometry of the complex), hence only the r_0 parameters were varied.

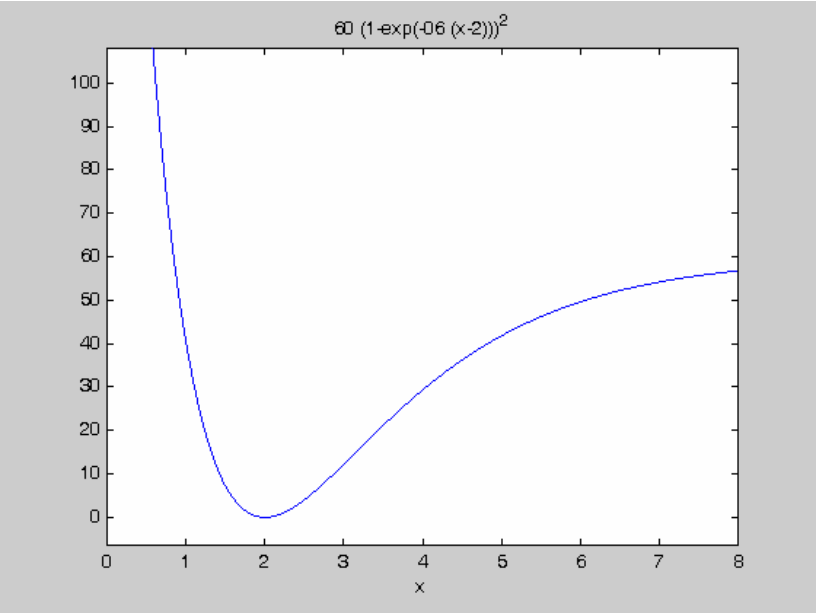


Figure 6-3: Morse function with parameters D = 60, α = 0.6, r_0 = 2.

The y-axis represents relative energy which is lowest at x=2 (the optimal bond length) but rises steeply at smaller bond lengths and less steeply at larger bond lengths. D is the depth of the energy "well" which is relative to the energy if the two atoms are infinitely apart.

6.5.3.2 : **Ligand-ligand repulsion terms**

The other DOMMIMOE parameters that needed to be set were the Fe – OM and Fe – O= ligand-ligand repulsion terms, A_{ll}. The ligand-ligand repulsion terms were important as these separate the oxygen atoms as far out as possible, which in a system where the ion is surrounded by six atoms these are arranged at the vertices of an octahedron. Large ligand-ligand repulsion terms increase the repulsion between the carbonyl and oxime oxygen atoms and will increase the Fe – oxygen distances. The parameters that were optimized simultaneously and the constants that were set beforehand are summarized in Table 6-5.

Table 6-5: DOMMIMOE parameters modified for ferric tris-hydroxamates.

Question marks represent unknown values or values that need to be found.

Parameter name	Description	Value set
Fe+3 – O=	Fe to carbonyl oxygen bond	
r_0	Optimal distance in Morse function	?
A_{ll}	Ligand-ligand repulsion term	?
α	Curvature parameter in Morse function	1.5
D	Depth parameter in Morse function	60
n	Ligand-ligand repulsion term	6
Fe+3 – OM	Fe to carbonyl oxygen bond	
r_0	Optimal distance in Morse function	?
A_{ll}	Ligand-ligand repulsion term	?
α	Curvature parameter in Morse function	1.5
D	Depth parameter in Morse function	60
n	Ligand-ligand repulsion term	6

6.5.3.3 : **Partial charges**

Another part of molecular model which has to be specified is its partial charges, in particular around the metal centre. The forcefield MMFF94 assigned partial charges between main group elements, which were not altered, but the charges around the Fe atom had to be specified. In metal complexes there is usually a highly charged metal cation surrounded by anionic or neutral ligands, and there is an issue in how much donation of electron density there is from the negatively charged ligands to the positively charged metal. Initially, a charge scheme based on natural charges given by the Jaguar quantum mechanical program[108] was used for the ferric-*tris*-hydroxamates, but suitable parameters could not be found by the parameter optimizing procedures explained later. Instead, the forcefield charges assigned by MMFF94 were used, with an extra bond charge increment term to decrease the positive charge on Fe(III) (Table 6-6).

Table 6-6: Charge schemes used in the parameterization of the MMFF94 forcefield of ferric-*tris*-hydroxamates in DOMMIMOE.

MMFF94 atom type	Description	Jaguar-based partial charge scheme	Bond charge increment scheme
Fe+3	Fe^{3+} ion	2.400	1.350
OM	oxime oxygen	-0.712	-0.512
O=	carbonyl oxygen	-0.670	-0.470
NC=O	oxime nitrogen atom in $Fe(hydr)_3$	-0.418	-0.468
NC=O	oxime nitrogen atom in $Fe(Nmhydr)_3$	-0.348	-0.398
C=	carbonyl atom	0.569	0.569
HN2	hydrogen atom attached to oxime nitrogen (only in $Fe(hydr)_3$)	0.370	0.370

6.5.4 : Parameter optimization by simulated annealing

In this study four parameters in DOMMIMOE were optimised: r_0 and A_{ll} for Fe – O=, and r_0 and A_{ll} for Fe – OM, which was sufficient to recover a suitable parameterization such that the DFT geometries of the *fac* isomers of $Fe(hydr)_3$ and

$Fe(Nmhydr)_3$ were reproduced accurately and also the energy difference between the absolute energies of these complexes were 5-6 kcal/mol less than their C_{3v} minimized geometries as in DFT. This was achieved by running a series of SVL scripts to find the parameters that optimized a cost function which was minimal when the MM geometries matched the DFT structures and the absolute energy difference between these geometries and their C_{3v} structures were close to 5 kcal/mol.

6.5.4.1 : **Comparison of structures obtained by DOMMIMOE and DFT**

From a given set of parameters it was necessary to obtain some measure of how the DOMMIMOE structures compared against the DFT structures. The most common indicators are the many bond lengths, internal bond angles, external bond angles and torsion angles needed to be matched, but this was difficult to implement. Another more straightforward way was to superimpose the structures at the Fe atom to obtain a measure of fit. This was straightforward to implement as MOE has an internal function called `Superpose` which did this process by finding the optimal superposition by minimizing the mean square deviation (MSD) between the atom positions in the two molecules. The `Superpose` function accepted two different structures of the same molecule and produced a smaller number if the structures fit better. A value of 0 from Superpose was indicative of a perfect match of geometries.

6.5.4.2 : **Optimizing C_{3v} structures in DOMMIMOE**

The optimization of the unsymmetrical structures by DFT in DOMMIMOE was obtained by using the `DOMMIMOE_Minimize` function. As DOMMIMOE had no inbuilt symmetry-restricted structure optimization of C_{3v} structures analogous to ADF, the `DOMMIMOE_Minimize` function was used again but with the Fe and surrounding oxygen atoms were fixed to stop the minimization to an “octahedral” structure. Due to the artificial “fixing” of seven atoms in the system in DOMMIMOE, the C_{3v} structures obtained was not compared to those obtained by DFT.

6.5.4.3 : **Structure of the cost function**

The two requirements of the cost function were that it was minimal when the minimal DOMMIMOE structures of *fac* $Fe(hydr)_3$ and *fac* $Fe(Nmhydr)_3$ matched the geometries obtained by DFT, and the absolute energy difference between the minimized and C_{3v} structures of *fac* $Fe(hydr)_3$ and *fac* $Fe(Nmhydr)_3$ in DOMMIMOE were the same as DFT. The structure of the cost function *C* is given in Equation 1. The optimization the DOMMIMOE parameters of simple ferric-*tris*-hydroxamates was reduced to minimizing the cost function *C*.

Equation 1: The structure of the cost function *C* used in parameter optimization.

***Fe(hydr)*$_3^{MOE}$ and *Fe(hydr)*$_3^{DFT}$ represent the minimal structures of $Fe(hydr)_3$ in DOMMIMOE and DFT respectively. *E(Fe(hydr)*$_3^{MOE}$*)* and *E(Fe(hydr)*$_3^{C3v}$*)* represent the absolute energy of the minimal and C_{3v} structures in DOMMIMOE in kcal mol^{-1} respectively. α and β are arbitrary constants.**

$$C(r_0^{Fe-O=}, A_{ll}^{Fe-O=}, r_0^{Fe-OM}, A_{ll}^{Fe-OM}) = \alpha C_1 + \beta C_2, \text{ where}$$

$$C_1 = \texttt{Superpose}[Fe(hydr)_3^{MOE}, Fe(hydr)_3^{DFT}] + \texttt{Superpose}[Fe(Nmhydr)_3^{MOE}, Fe(Nmhydr)_3^{DFT}]$$

$$C_2 = \text{Abs}[E(Fe(hydr)_3)^{MOE}, E(Fe(hydr)_3)^{C3v}] + \text{Abs}[E(Fe(Nmhydr)_3)^{MOE}, E(Fe(Nmhydr)_3)^{C3v}]]$$

6.5.5 : MCMC and simulated annealing

There are many different methods and algorithms which could be used to minimize a cost function, whose suitability depends on a number of factors. Simulated annealing is commonly used class of algorithms to find suitable parameters to minimize a cost function; especially suitable if the parameter space contains many local minima or has an unknown distribution. It uses the results of Markov Chain Monte Carlo (MCMC) which is used by statisticians and other disciplines to find out information about unknown distributions.

6.5.5.1 : **Theory of MCMC**

MCMC is a way of sampling an unknown distribution $\pi(x)$ by performing a "random walk" on it. Strictly, the distribution should be a probability density function whose area or sum is identical to one, but the results are applicable to positive distributions

(i.e. $\pi(x) > 0$). In MCMC[109], one constructs an ergodic Markov Chain that has a limiting distribution $\pi(x)$. In the case of simulated annealing, this is done by an algorithm called the Metropolis-Hastings sampler[109], whose principle is summarized below.

1. Choose a starting point *x*
2. Choose a next state y from a **proposal** distribution q(*x*,*y*), so-called as it suggests the next step in the Markov Chain. q can be any distribution, but in the case of simulated annealing and many other MCMC algorithms this is a uniform distribution.
3. Accept y as the next step with probability $\alpha(x,y) = \min\{1, \frac{\pi(x)}{\pi(y)}\}$ if q(*x*,*y*) = q(*y*,*x*). This is known as the **Metropolis** criteria. The interpretation of this is: if $y<x$ then accept the new step, otherwise accept the step with probability $\pi(x)/\pi(y)$.
4. If *y* is accepted, then let *y* be the next step of the Markov Chain. Otherwise let *x* be the next step.
5. Repeat steps 1-4

6.5.5.2 : **Simulated annealing**

In simulated annealing, the parameter space is explored by an ergodic Markov chain which has a tendency to move 'downhill'. To minimize the cost function $C(x)$, starting from *x*, another move *x'* is proposed from a uniform distribution U[$x - VM$, $x + VM$]. *VM* is the maximum stepsize, and is adjusted such that the steps in the Markov chain are not too big or small. If $C(x') < C(x)$ then let *x'* be the next step, otherwise accept *x'* to be the next step with probability $\left(\frac{C(x)}{C(x')}\right)^{1/T}$ where $0 < T \leq 1$. Downhill moves are always accepted; uphill moves can be accepted, but less so the larger the uphill step proposed is.

The temperature factor *T* is the key concept in simulated annealing. At the start of the optimisation *T* is set to 1 (the Markov chain is trained on $C(x)$ initially), but decreases as it progresses. Before a change of *T*, the algorithm starts from *x* such

that $C(x)$ was minimal. A smaller T reduces the chance of accepting larger steps, and as T decreases the calculation is hoped to settle at the global minimum (Figure 6-4).

A FORTRAN-based simulated annealing algorithm[110] was translated into SVL to minimize the penalty function $C(r_0^{Fe-O=}, A_{ll}^{Fe-O=}, r_0^{Fe-OM}, A_{ll}^{Fe-OM})$ within the bounds given in Table 6-7. When using the charge scheme based on the natural charges suggested by the Jaguar quantum mechanical program, the energy difference between the octahedral and C_{3v} structures were too high in DOMMIMOE compared to the DFT difference although the geometry fits were almost perfect. Attempts to find parameters which reproduced the energy difference involved introducing α and D terms into the penalty function in the simulated annealing to increase the flexibility of the parameter search. Eventually, the problem was solved by using the MMFF94 charge scheme with bond-charge increments.

Table 6-7: Bounds of the simulated annealing parameter optimization.

Parameter name	Bounds
Fe+3 – O=	
r_0	1 < x < 3
A_{ll}	500 < x < 20000
Fe+3 – OM	
r_0	1 < x < 3
A_{ll}	500 < x < 20000

6.5.6 : Results of simulated annealing

After the optimization (whose parameters are summarized in Table 6-8), the geometries and energy differences of DOMMIMOE compared to the DFT calculations were good (Table 6-9). The slightly higher RMSD value for $Fe(Nmhydr)_3$ compared to $Fe(hydr)_3$ was due to the rotation of the three hydrogen atoms on the extra methyl groups on the nitrogen due to the forcefield.

Table 6-8: Table of parameters found by simulated annealing applied on a penalty function given by Equation 1.

Parameter name	Value
Fe+3 – O=	
r_0	2.050
A_{ll}	2504
Fe+3 – OM	
r_0	1.986
A_{ll}	1505

Table 6-9: The performance of the parameters found by simulated annealing.

Complex	MSD (RMSD) of DOMMIMOE from DFT structure (Å)	DOMMIMOE Energy difference from C_{3v} structure kcal mol^{-1}	DFT Energy difference from C_{3v} structure kcal mol^{-1}
$Fe(hydr)_3$	0.03341 (0.1828)	6.358	6.29
$Fe(Nmhydr)_3$	0.23918 (0.4891)	5.962	5.83

6.5.7 : Parameter Refinement

Although DFT structures compared very well with X-ray crystallographic structures, they were not exact. For example, DFT overestimated the Fe – O bond lengths in both facial and meridional isomers of $Fe(hydr)_3$ by 0.06 Å from the X-ray crystallographic structure by Failes and Hambley[71]. Hence, a parameter refinement was performed against 23 X-ray crystallographic structures of mononuclear ferric tris-hydroxamates (given in Appendix table 2). This required construction of a new cost function C'(*x*), which was the MSD between the X-ray structures and their DOMMIMOE minimized counterparts, calculated again using the internal `Superpose` function that calculates the RMSD between two structures (Equation 2). As a local optimization of parameters was required, Nelder-Mead simplex optimization was performed instead of simulated annealing, which was faster.

Equation 2: Cost function *C'* for parameter refinement by simplex optimization.

The terms *structure*$^{\text{X-ray}}$ and *structure*$^{\text{MOE}}$ represent the X-ray structure in the database and its DOMMIMOE minimized structure respectively.

$$C'(x) = \sum_{structure=1}^{24} \mathrm{Superpose}[structure^{\mathrm{X-ray}}, structure^{\mathrm{MOE}}]$$

6.5.7.1 : **Nelder-Mead simplex optimization**

The Nelder-Mead simplex method[111] is a commonly used nonlinear optimization algorithm for minimising an objective function in a many-dimensional space. It is a rather robust algorithm and is rather fast at converging to minima, but does not explore the parameter space very effectively, hence is likely to converge to local minima. As only a little parameter refinement is needed and evaluating the new cost function takes approximately 2 minutes (from the time take to minimize 23 ferric tris-hydroxamate structures in DOMMIMOE) compared to 5 seconds with the previous cost function, a method which explored local parameter space plus did not require as many penalty function evaluations as simulated annealing .

The algorithm starts from a point $\boldsymbol{x_1}$ in n-dimensional parameter space. Then a simplex of n+1 vertices in n dimensions is constructed, for example, if $\boldsymbol{x_1}$ is two dimensional then two other points are chosen nearby $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$. In the two-dimensional case the simplex is a triangle. The cost function is evaluated at these points and depending on its values the triangle is reflected, extended or contracted until a minimal or maximal triangle of points is constructed, depending if the function needs to be minimized or maximized respectively. The optimization on the cost function $C'(r_0^{Fe-O=}, A_{ll}^{Fe-O=}, r_0^{Fe-OM}, A_{ll}^{Fe-OM})$ was a four-dimensional problem hence a five-dimensional simplex was constructed.

The simplest step is to replace the worst point with a point reflected through the remaining N-1 points considered as a plane. If this point is better than the best current point, then the algorithm will try to stretch out exponentially along this line. On the other hand, if this new point is not much better than the previous value then

the algorithm shrinks the simplex towards the best point, as there is a 'valley' between the worst and new points which required exploration (Figure 6-5). The algorithm was translated into SVL from the MATLAB code of the `fminsearch.m` function.

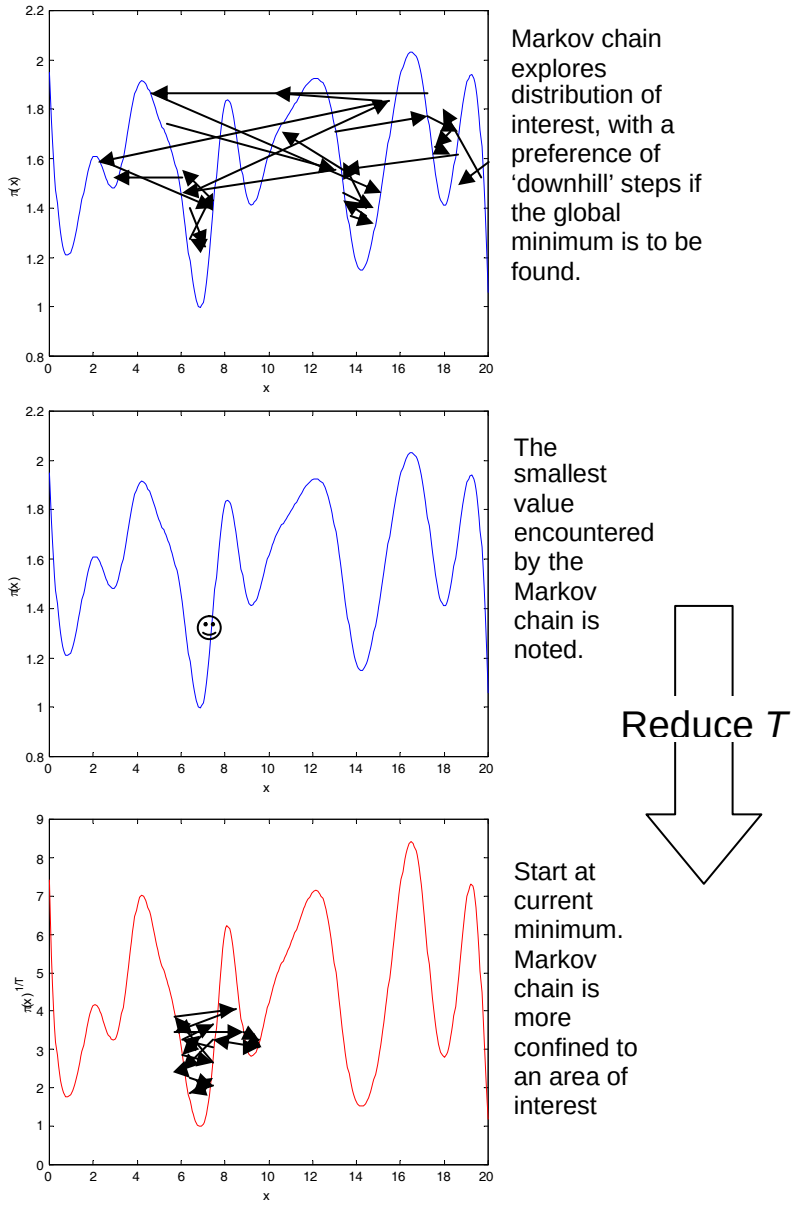


Figure 6-4: The concept of simulated annealing.

The curve represents an unknown distribution such as a cost function with multiple minima. A Markov chain is trained onto the curve to explore the state space. After multiple iterations, the Markov chain is started onto the minimum encountered so far, but with the simulated annealing temperature *T* reduced, which has the effect of making the state space 'steeper', keeping the chain around the minima. This process is repeated until a minima is reached, which hopefully will be the global minimum.

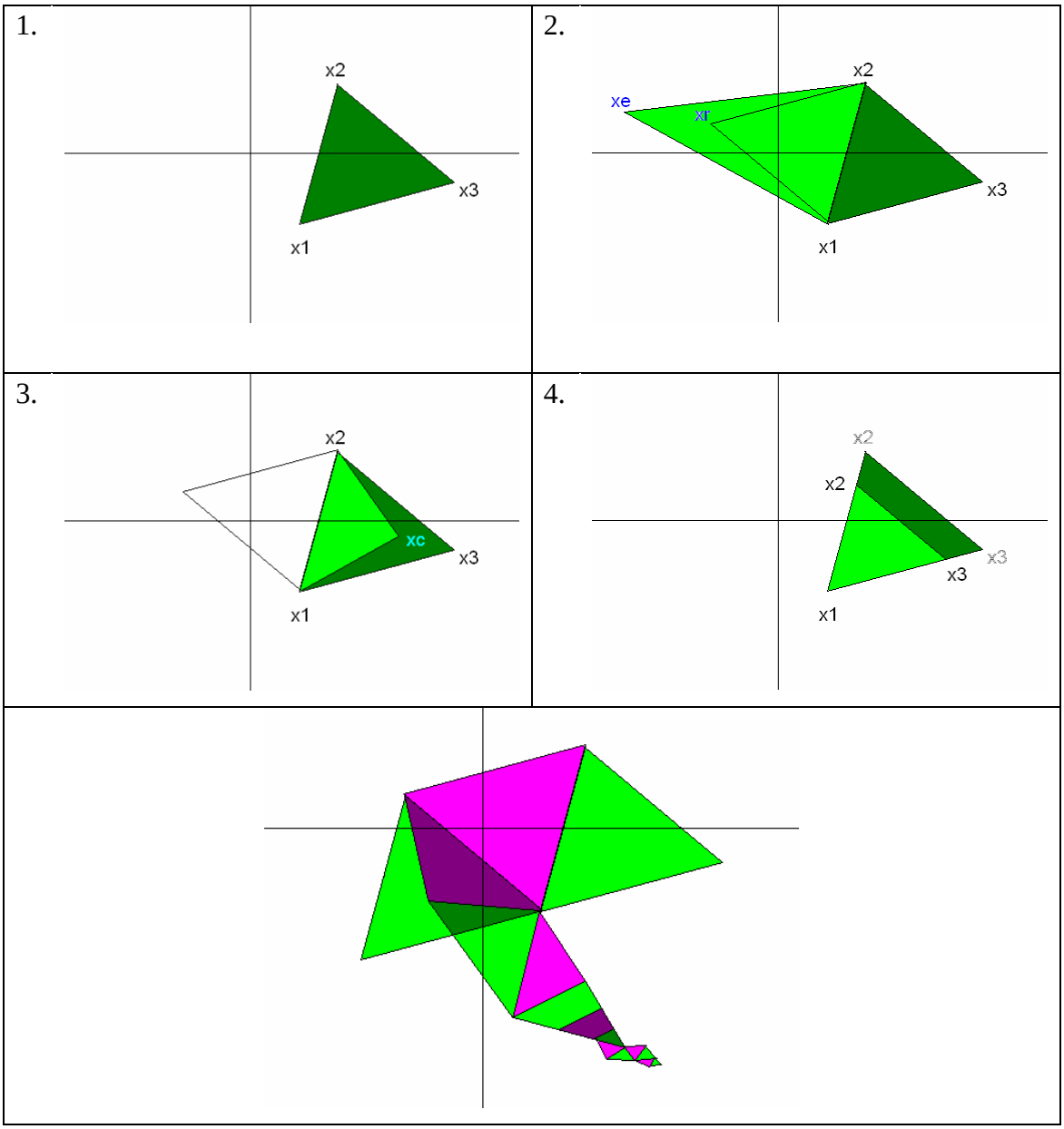


Figure 6-5: Simplex optimization in two dimensions.

In this example, parameter space is in two dimensions. (1) The initial simplex is a triangle, where x_1, x_2 and x_3 represent initial guesses to the minimum of the cost function C(x). For instance if $C(x_1) \leq C(x_2) \leq C(x_3)$ i.e. x_3 is the worst point, then a reflection of x_3 off the line x_1x_2 is computed, x_r. (2) If $C(x_r) < C(x_1)$ i.e. x_r is the current minimum then an extension along the line x_3x_r is computed. If $C(x_e) < C(x_r)$ make new simplex with x_3 replaced by x_e otherwise replace with x_r. (3) If $C(x_r) > C(x_2)$ then a contraction is performed towards the line x_1x_2. If $C(x_c) < C(x_3)$ then reject x_3 for x_c in the new simplex. (4) If $C(x_r) > C(x_3)$ i.e. x_r is the worst point then shrink the simplex to x_1. The process is repeated on the new simplex until it is sufficiently small. (Bottom) An example path to a minimum by the Nelder-Mead simplex algorithm.

6.5.7.2 : **Results of Parameter Refinement**

The parameter search took approximately 36 hours on the 3.0 GHz system as described in the Experimental section, and are given in Table 6-10.

Table 6-10: Table of parameters found after parameter refinement via 23 X-ray crystal structures of the parameters found by simulated annealing

Parameter name	Parameter found
Fe+3 – O=	
r_0	1.9860
A_{ll}	2151
Fe+3 – OM	
r_0	1.9251
A_{ll}	1375

To demonstrate the performance of these parameters on the modelling of ferric-*tris*-hydroxamates, the bond lengths, bite angles and trigonal twist RMSDs around the metal centre against the 23 X-ray crystal structures in the database were calculated. This is summarized in Appendix Table 0-3.

6.6 : Performance of parameters

6.6.1 : Fe – O bond lengths

The Fe – O (carbonyl) bond lengths were overestimated by on average 0.0011 Å by DOMMIMOE compared to the X-ray crystal structures, with standard deviation of 0.016 Å. The root mean squared deviation (RMSD) of these bond lengths was 0.017 Å on average. The Fe – O (oxime) bond lengths were underestimated on average 0.0018 Å by DOMMIMOE, with standard deviation 0.016 Å. The RMSD of these bond lengths was 0.017 Å on average.

6.6.2 : O – Fe – O bite angles in the same hydroxamate

The O(carbonyl) – Fe – O(oxime) bite angles were overestimated by 0.02° by DOMMIMOE compared to the X-ray crystal structures, with standard deviation 0.8°. The RMSD of these bond angles was 0.7° on average.

6.6.3 : Trigonal twists

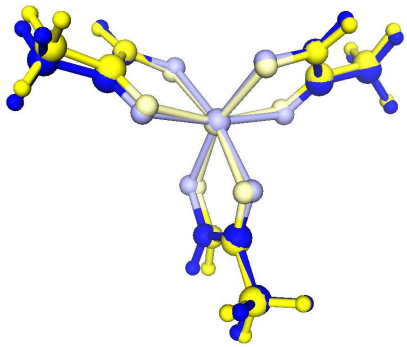
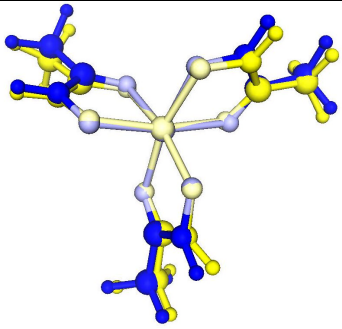
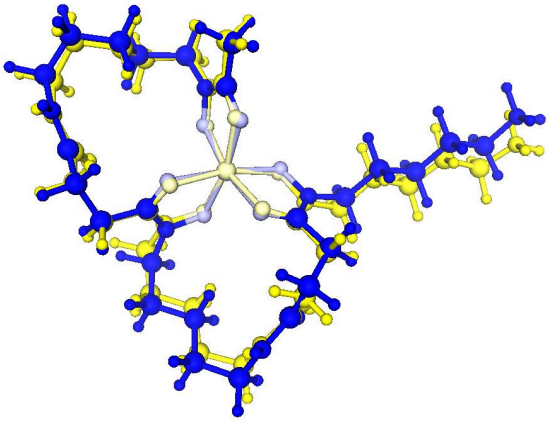
The trigonal twist (see section 6.3.1) was the main distortion amongst ferric-*tris*-hydroxamates. The parameterization of DOMMIMOE overestimated the trigonal twists of the crystal structures on average by 5.4°, with standard deviation 3.8°, whose interpretation is that DOMMIMOE makes the ferric complex more "octahedral" than the X-ray crystallographic structure. There was much local variation: the trigonal twist difference between the DOMMIMOE and X-ray crystal structure of *fac* ferric-*tris*-acetohydroxamic acid[71] (SUXREI) and ferrioxamine E[104] (FEROXE10) was only 3.7° and 4.3° respectively, whilst ferricrocin[112] (FERCRN10) and tris(Benzohydroxamato-O,O')-iron(iii)[71] (FEBOAH02) had a difference of 5.7° and 12.2° respectively. The trigonal twists of the *mer* complexes COFDIK10[113] (neocoprogen I) and meridional ferric-*tris*-acetohydroxamic acid were overestimated on average 0.7° and 9.3° respectively. The RMSD of the twist angles was 6.6° on average.

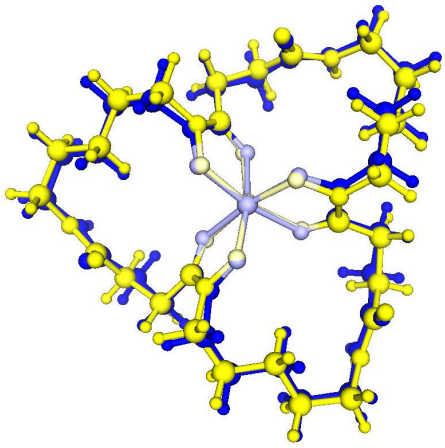
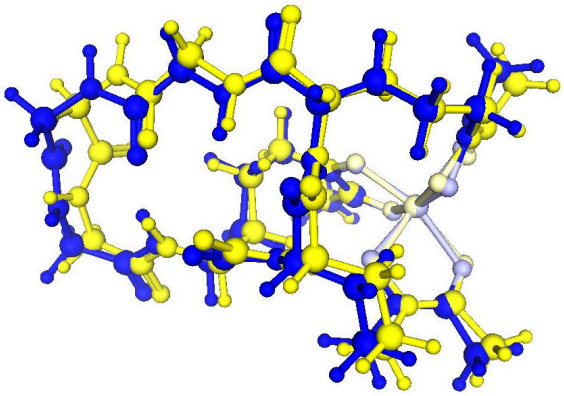
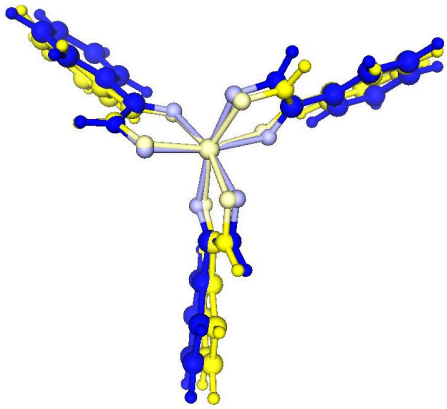
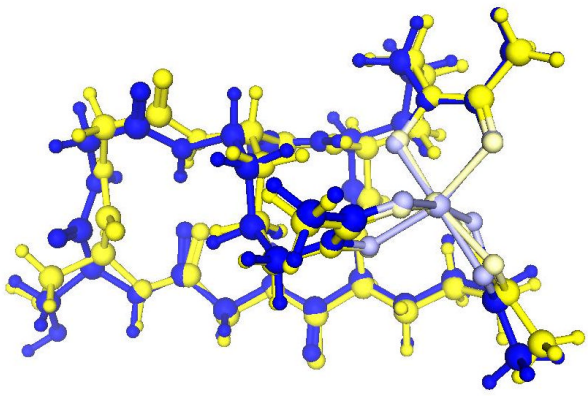
6.6.4 : Superposition of DOMMIMOE and X-ray structures

To demonstrate the performance of the DOMMIMOE parameters, the minimized structures were superimposed onto the original X-ray crystal structures, with the Fe atom set at the centre (Table 6-11). The superpositions of *fac*- and *mer*- $Fe(hydr)_3$, ferrioxamines E and B and ferrichrome show a very good correspondence between the DOMMIMOE and crystal structures, and even in ferricrocin (FERCRN10) and tris(Benzohydroxamato-O,O')-iron(iii) (FEBOAH02) where the discrepancy between the trigonal twists are large, the structures show reasonable correspondence.

Table 6-11: Superposition of seven X-ray ferric tris-hydroxamate structures with the DOMMIMOE optimized structures, with the superposition centred on the Fe(III) ion and its surrounding oxygen atoms.

Yellow – X-ray structures, Blue – DOMMIMOE optimization; lighter colours – Fe and neighbouring oxygen atoms.

Name and CCSD Refcode	DOMMIMOE/ X-ray structure superposition
SUXREI (*fac*- Fe(hydr)$_3$)	
SUXREI (*mer*- Fe(hydr)$_3$)	
UNEWAL (Ferrioxamine B)	

FEROXE10 (Ferrioxamine E)	
FERRID10 (Ferrichrome)	
FEBOAH02 tris(Benzohydroxamato-O,O')-iron(iii)	
FERCRN10 (Ferricrocin)	

6.7 : Geometry scans

6.7.1 : Stretch strain energies of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ in DOMMIMOE compared to DFT

To test whether stretching distortions could be modelled by DOMMIMOE, the ligands of the complexes $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ - in particular the centroids of the hydroxamate groups - were systematically shifted away or towards the Fe atom in 0.05 Å increments (Figure 6-6), and the single point energy (without minimization) was taken in DOMMIMOE and in ADF. The absolute energy of the minimal complex was subtracted from the single point energies and graphed to illustrate a "stretching strain energy" as the ligands move towards or away from the minimal structure (Figure 6-7). The curves show excellent agreement.

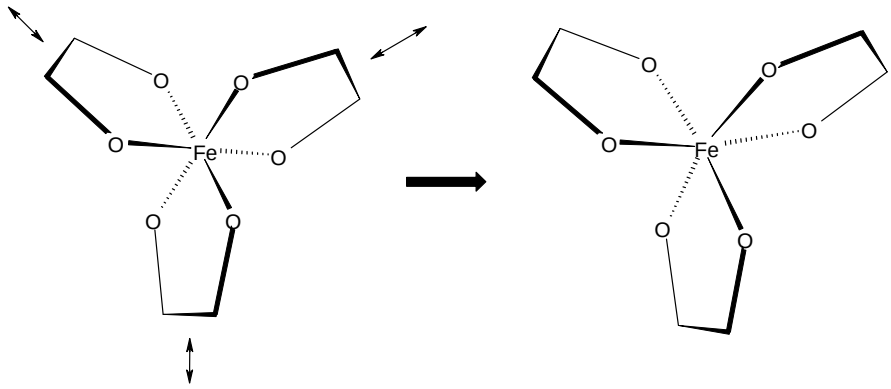


Figure 6-6: Diagram showing the adjustment of the hydroxamate ligands in the measuring of the stretch potentials.

The ligands were systematically moved towards or away from the Fe atom by moving the centroid of the hydroxamates.

6.7.2 : Trigonal distortion strain of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ in DOMMIMOE compared to DFT

To test whether the energies required for trigonal distortion could be modelled by DOMMIMOE, geometries of the complexes of the facial isomers of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ were constructed where their trigonal twists approached zero (Figure 6-8). This was achieved by rotating the hydroxamate ligands about their Fe – centroid axes in increments of 0.05 radians to 0.6 radians, and their single point energies recorded and subtracted from the minimal geometry energy to calculate a

"trigonal distortion strain energy". A similar process was also carried out in DFT, using the DFT minimized $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ as a starting point (Figure 6-9).

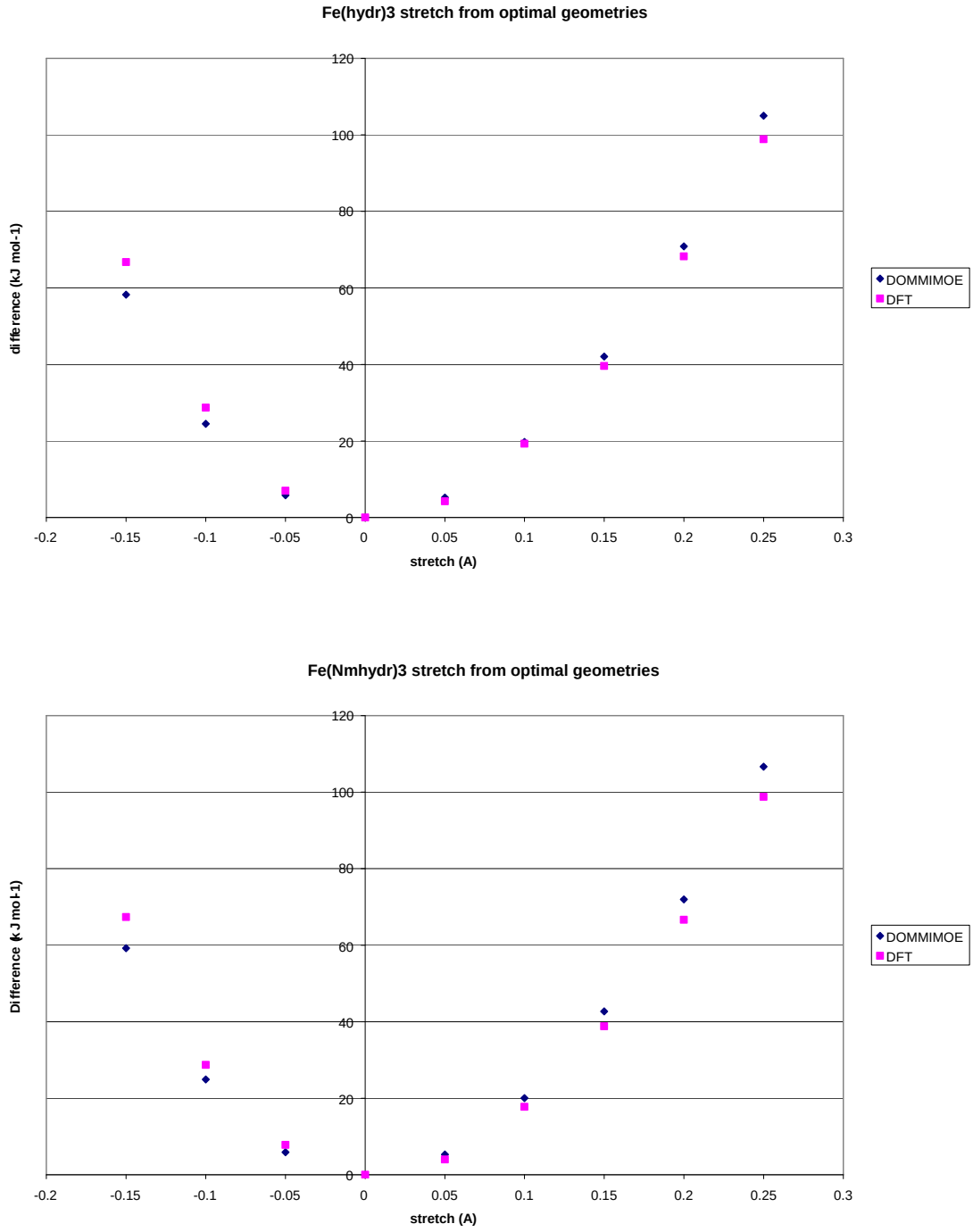


Figure 6-7: The difference in absolute DFT/ DOMMIMOE energies from optimal geometries when each hydroxamate ligand is extended simultaneously

(Top) $Fe(hydr)_3$; (Bottom) $Fe(Nmhydr)_3$

For $Fe(hydr)_3$, the DOMMIMOE energy differences of these trigonally distorted complexes from the minimal geometry closely matched those from DFT calculation, with the difference between optimal and trigonal prismatic geometry (ligand rotation 0.55 radians; trigonal twist 2°) for both DOMMIMOE and DFT being 31 kJ mol^{-1} (7.3 kcal mol^{-1}). The offset between the maximum of the DOMMIMOE and the DFT curves comes from the fact that the optimal geometry of the DFT structure (trigonal twist 43°) was more trigonally distorted than the DOMMIMOE structure (trigonal twist 48°), hence less rotation of the ligands was required to achieve greatest energy difference at trigonal twist close to zero.

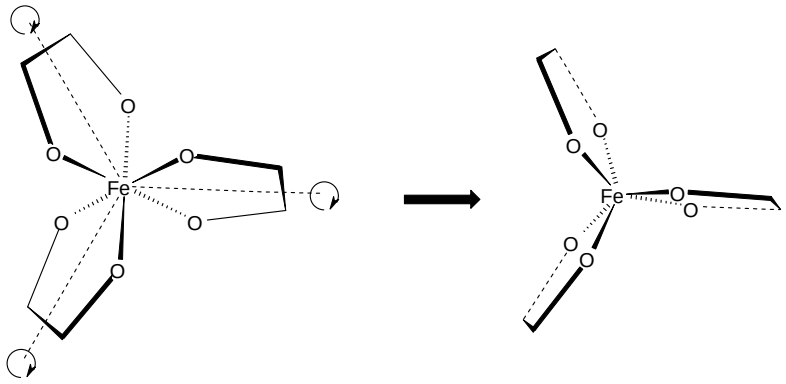


Figure 6-8: Diagram showing the adjustment of the hydroxamate ligands in the measuring of the trigonal twist potentials.

The ligands were rotated about the Fe-centroid axes towards a trigonal prismatic geometry.

This was also the case for $Fe(Nmhydr)_3$ up to a rotation of 0.3 radians. This corresponded to a trigonal twist of 24.6°, lower than the trigonal twists observed in the X-ray crystal structures. As the trigonal twist of $Fe(Nmhydr)_3$ tended to zero, the difference between DFT and the DOMMIMOE energies from optimal geometries increased to 5.67 kJ mol^{-1} (1.4 kcal mol^{-1}), hence the difference between the optimal and trigonal prismatic geometry for DOMMIMOE and DFT was about 31.9 kJ mol^{-1} and 25.4 kJ mol^{-1} respectively (a difference of 6.5 kJ mol^{-1}).

In consideration of using this model in molecular dynamics simulations, it is important to note what energy barriers are needed to be overcome to form the twisted structures of ferrioxamine B found bound to siderophore-binding proteins such as *S.*

coelicolor DesE and *E. coli* FhuD. This difference between the energies of 6.5 kJ mol^{-1} is relatively insignificant.

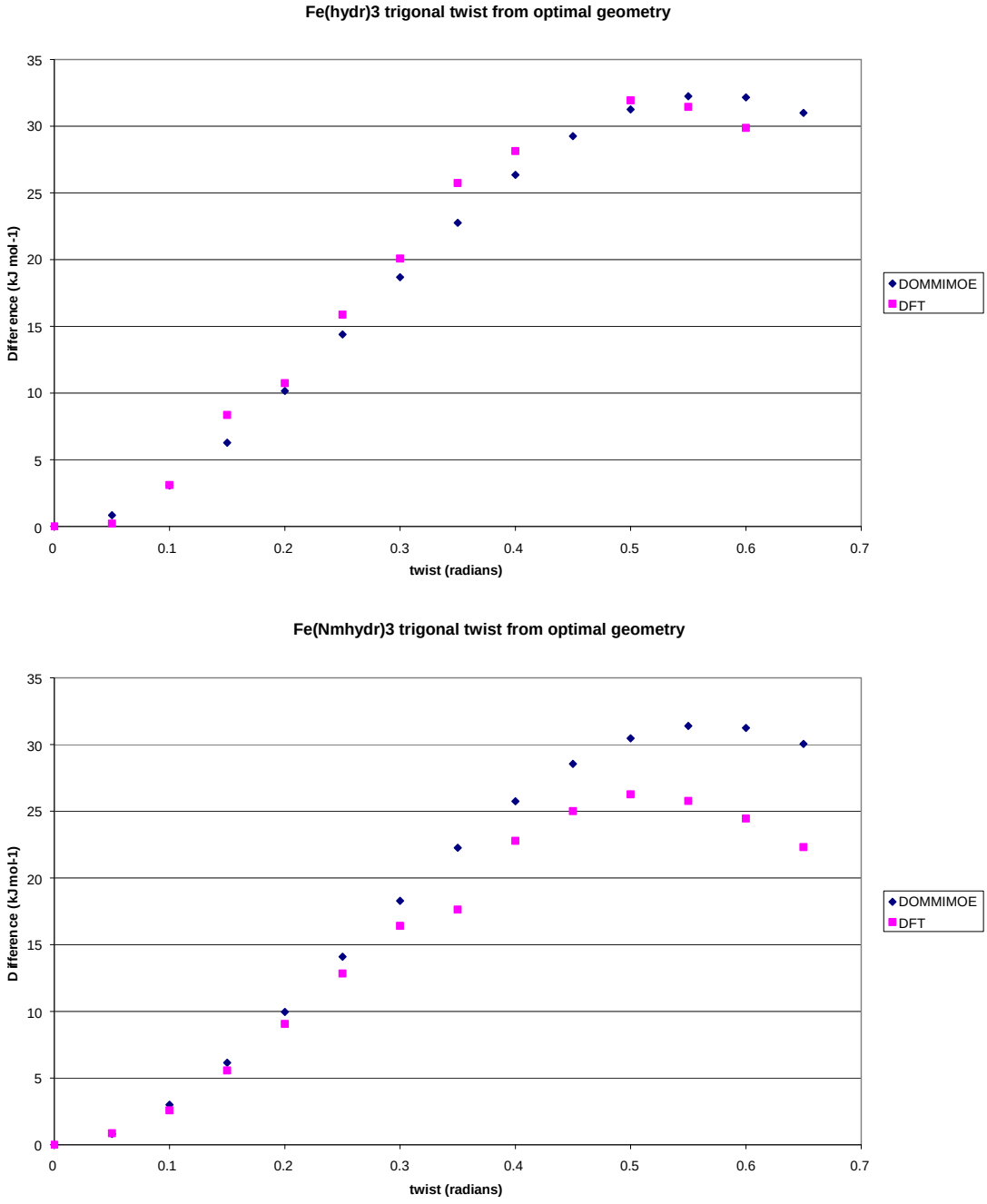


Figure 6-9: The difference in absolute DFT/ DOMMIMOE energies from optimal geometry when each hydroxamate ligand twisted simultaneously about their centroids towards a trigonal prismatic geometry.

(Top) $Fe(hydr)_3$; (Bottom) $Fe(Nmhydr)_3$.

6.8 : Comparison with other models

The Fe – O bond lengths and O – Fe – O bite angles obtained by this parameterization for $Fe(hydr)_3$ performed better than the study by Zinelabidine *et al.*[103] as demonstrated by Table 6-12. The forcefield produced was trained to replicate the crystal structure of ferrioxamine E as there was no structure available for $Fe(hydr)_3$. For instance, the Fe – O carbonyl distance was on average 0.048 Å less than those given in X-ray crystal structures compared to 0.011 Å in DOMMIMOE. Also the bite angle O(N) – Fe – O(C) was 3.4° greater than the reference X-ray crystal structure, where the same figure is 0.2° for DOMMIMOE.

The model constructed by Herrera *et al.*[105] was even worse at reproducing the geometries of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ (Table 6-13), with Fe – O(carbonyl) bonds underestimated by 0.18 Å, and the bite angle overestimated by 7.8° and 11° in the case of $Fe(hydr)_3$ and $Fe(Nmhydr)_3$ respectively, implying that the oxygen atoms around the central Fe(III) ion were almost octahedral, which is not the case in the X-ray crystal structures of $Fe(hydr)_3$ or ferrioxamine E.

Table 6-12: Comparison of the Fe – O bond lengths and angles of the crystal structure of $Fe(hydr)_3$, ferrioxamine E, and $Fe(hydr)_3$ optimised with DOMMIMOE and the method of Zinelabidine et al[103]

	X-ray $Fe(hydr)_3$ [71]	DOMMIMOE $Fe(hydr)_3$	X-ray (Ferrioxamine E)[104]	Zinelabidine *et al.* [103] $Fe(hydr)_3$
Average Fe- O(C) bond length (Å)	2.056	2.045	2.055	2.007
Average Fe-O(N) bond length (Å)	1.970	1.984	1.952	1.941
Average O(C)–Fe–O(N) angle (°)	79.7	79.5	78.9	82.3

Another test of the model of Herrera *et al.* was measuring the energy difference of *fac* and *mer* isomers of $Fe(Nmhydr)_3$. In DFT, the energy difference was 4.0 kcal mol^{-1}, with the *mer* isomer being higher in energy, whilst their model gave a value of 3.6 kcal mol^{-1}, which was surprisingly close to the DFT value. The DOMMIMOE model gave a value of 0.7 kcal mol^{-1}, with the *fac* isomer being higher in energy, in

contrast to the DFT results. In this respect the DOMMIMOE model performed less well than the model of Herrera *et al.*

Table 6-13: Comparison of the Fe – O bond lengths and angles of the crystal structure of $Fe(hydr)_3$, DFT structure of $Fe(Nmhydr)_3$, and $Fe(hydr)_3$ optimised with DOMMIMOE and the method of Herrera *et al.*

$Fe(hydr)_3$			
	X-ray [71]	**DOMMIMOE**	**Herrera *et al.*[105]**
Average Fe-O(C) bond length (Å)	2.056	2.045	1.880
Average Fe-O(N) bond length (Å)	1.970	1.984	1.874
Average O(C)-Fe-O(N) angle (°)	79.7	79.5	87.5
$Fe(Nmhydr)_3$			
	DFT	**DOMMIMOE**	**Herrera *et al.*[105]**
Average Fe-O(C) bond length (Å)	2.10	2.042	1.875
Average Fe-O(N) bond length (Å)	2.00	1.977	1.841
Average O(C)-Fe-O(N) angle (°)	77.6	79.2	88.5

6.9 : Future work

The molecular model parameterization found would be ideal to predict the structure of novel siderophores. A similar parameterization for Ga(III)-hydroxamates was used to solve the structure of the novel siderophore coelichelin in complex with Ga(III)[52]. This was achieved by the construction of isomers of the tetrapeptide coelichelin with Ga(III) in DOMMIMOE, and their interatomic distances were compared with those calculated by two-dimensional NMR data of gallium-coelichelin.

The molecular dynamics function developed recently would be ideal to model the interaction of the *S. coelicolor* protein DesE with different ferri-siderophores, with

the recently solved structure of the DesE-ferrioxamine B complex could be used as a starting point. Ferrioxamine B was found in a "twisted" conformation in complex with DesE and also with *E. coli* FhuD (although this protein is not a strong binder to this siderophore) unlike that found in ferrioxamine B crystallized without proteins[72]. The flexibility of siderophores has successfully been taken into account by consideration of the trigonal twist.

The DOMMIMOE model could be used to predict the binding of this protein with other ferri-siderophores. The ferrioxamine B in the crystal structure could be replaced by models of ferri-siderophores tested by the fluorescence quenching assay to determine if important electrostatic and hydrogen bonding networks between the amino acid residues and siderophore could be identified, plus whether the dissociation constant of the receptor-siderophore complex could be estimated computationally using molecular dynamics in DOMMIMOE.

The other two proteins implicated in siderophore binding in *S. coelicolor,* CchF and CdtB, are sufficiently similar to other crystallized siderophore periplasmic binding proteins to produce homology models. CchF and CdtB have an amino acid identity of 27%, with 44% positive alignments with *S. coelicolor* DesE and *Bacillus subtilis* FeuA[34] respectively. These models could be used as suggested with DesE to find the main factors that lead to their selective binding of ferric-siderophores.

6.10 : Conclusions

A molecular model of ferric-*tris*-hydroxamates has been constructed in DOMMIMOE, which accurately reproduced bond lengths and angles of X-ray crystal structures around the Fe(III) centre, but slightly overestimated the trigonal distortion in these complexes. A range of techniques have been used to find the best parameterization, such as the utilization of structural and energetic information from DFT calculations, SVL programming, simulated annealing and Nelder-Mead simplex optimization. The stretching and twisting of ferric hydroxamate ligands in model complexes have shown this model may be able to model the flexibility of these complexes, and would be more than adequate to use in molecular dynamics

simulations, in particular to model the binding of ferric hydroxamate siderophores to siderophore binding proteins.

Chapter 7: Summary, conclusions and future work

7.1 : Summary and Conclusions

7.1.1 : Purification and characterization of three siderophore-binding receptor proteins of *Streptomyces coelicolor* A3(2)

The main aims of this study were to overproduce and purify the three siderophore-binding receptor proteins DesE, CchF and CdtB of *Streptomyces coelicolor* A3(2) in *Escherichia coli*, and to assess their ability to bind to cognate ferri-siderophores by monitoring the decrease of the fluorescence of the proteins upon ligand binding.

Initially, the full-length DesE, CchF and CdtB proteins were overproduced in *E. coli* with a hexahistidine tag appended to their native N-termini. All three proteins were overproduced in a soluble form, and two were purified by immobilized metal-affinity chromatography (IMAC). CchF, whose His-tag was most probably cleaved by the TAT translocase machinery of *E. coli*, could not be purified by this method. The other proteins, DesE and CdtB, dimerized very strongly, as indicated by denaturing SDS-PAGE and size-exclusion chromatography (gel filtration).

New plasmids were constructed for the expression of 5'-truncated *desE, cchF* and *cdtB* genes in *E. coli.* These encoded the recombinant proteins DesE, CchF and CdtB without the hydrophobic N-terminal signal sequences and lipidation sites. All three proteins were overproduced in soluble form and purified by Ni-IMAC. DesE and CdtB were monomeric, whereas CchF was found to dimerize (using gel filtration). This CchF dimer was dissociated by incubation at 37 °C. The N-terminal His-tags were removed from the proteins using tobacco etch virus (TEV) protease. The identity of the purified proteins was confirmed by peptide mass fingerprinting and ESI-MS analysis.

Circular dichroism analysis of the three proteins showed that two of them, DesE and CchF, were fully folded, whilst CdtB was up to 25% unfolded. This may explain the larger apparent molecular weight (corresponding to a larger hydrodynamic radius) of this protein indicated by gel filtration than the more massive proteins DesE and CchF.

7.1.2 : Intrinsic fluorescence quenching of proteins by ferri-siderophore binding

Solutions of ferri-siderophores were prepared and their concentrations were determined by inductively coupled plasma optical emission spectroscopy (ICP-OES). Ferri-coelichelin was purified from iron-deficient cultures from *Streptomyces coelicolor* using reverse-phase HPLC, utilising a modification of the method of Lautru et al[52], to produce gallium-coelichelin, whose purity was examined by ^{1}H NMR spectroscopy. Ferri-coelichelin was made by adding an excess of $FeCl_3$ to gallium-coelichelin.

Solutions of the proteins DesE, CchF and CdtB showed significant decreases in their intrinsic fluorescence upon addition of low concentrations of certain ferric siderophores. Dissociation constants were calculated from the fluorescence quenching data by fitting to a one-site binding curve using the equations of Ahnstrom et al[85] in the software package ORIGIN[86].

DesE was found to bind the cognate siderophores ferrioxamine B and ferrioxamine E with dissociation constants in the nanomolar range, but did not bind ferri-coelichelin with nanomolar affinity. DesE also bound the hydroxamate xenosiderophore ferrichrome with nanomolar affinity, but ferri-albomycin bound more weakly. CchF bound ferri-coelichelin only, whilst CdtB was found to bind ferrioxamine B and ferri-coelichelin with nanomolar affinity. None of the proteins studied bound the *E. coli* siderophore ferri-enterobactin, nor did the desferri-siderophores enterobactin, desferri-ferrichrome, and desferrioxamines E and B.

When CdtB was left to stand at room temperature in solution, the fluorescence decreased steadily, likely to be due to its instability. Despite this, the dissociation constants of ferrioxamine B and ferri-coelichelin from CdtB were in the nanomolar range even when it was assumed that only 75% of the protein was folded.

7.1.3 : X-ray crystal structures of DesE and DesE-ferrioxamine B

The X-ray crystal structures of apo-DesE and DesE-ferrioxamine B were determined by our collaborators, at the University of St. Andrews. The co-complex crystal structure confirmed the specific nature of the binding of ferrioxamine B to DesE, as suggested by the fluorescence quenching data. Also, little conformational change was observed in DesE on substrate binding, which was also seen in other siderophore binding proteins.

Analysis of the DesE-ferrioxamine B structure found several aromatic residues around the ferri-siderophore binding site, explaining the large quenching of the intrinsic fluorescence of DesE upon ferrioxamine B binding. An important hydrogen-bonding interaction of an arginine residue with the hydroxamate groups surrounding the ferric iron centre of ferrioxamine B was identified. Also, ferrioxamine B adopted an unusual configuration never observed before in X-ray crystal structures of this ferri-siderophore.

7.1.4 : Construction of a molecular model of ferric-tris-hydroxamates

Another aim of the project was to construct a molecular model for ferric-*tris*-hydroxamate complexes (including ferric-*tris*-hydroxamate siderophores) using the molecular mechanics program DOMMIMOE. This involved parameterization of the program such that it reproduced the geometries observed in X-ray crystallographic structures of ferric-*tris*-hydroxamates, with energetic information from quantum mechanical calculations on simple ferric-*tris*-hydroxamates to incorporate the trigonal twist, the main distortion of these complexes seen in the crystal structures.

A suitable parameterization was found by applying the techniques simulated annealing and Nelder-Mead simplex optimization to a cost function whose value was minimal if the geometries and energies produced by the program were close to the desired values.

The bond lengths around the metal centre observed in the X-ray crystal structures were accurately reproduced, as well as the ligand bite angle. The estimation of trigonal twist in X-ray crystallographic structures was less successful, with the

molecular model overestimating this by on average 5.4°. The model was more successful in predicting the quantum mechanical energies required to deform simple ferric-*tris*-hydroxamates by stretching and rotation of the ligands – the latter to imitate trigonal distortion of the complexes.

7.2 : Conclusions and future work

7.2.1 : Revised model for cognate siderophore uptake in *Streptomyces coelicolor* A3(2)

From the results of this study and the results of Barona-Gómez et al.[53], a revised model of ferri-siderophore uptake in *S. coelicolor* is proposed in Figure 7-1.

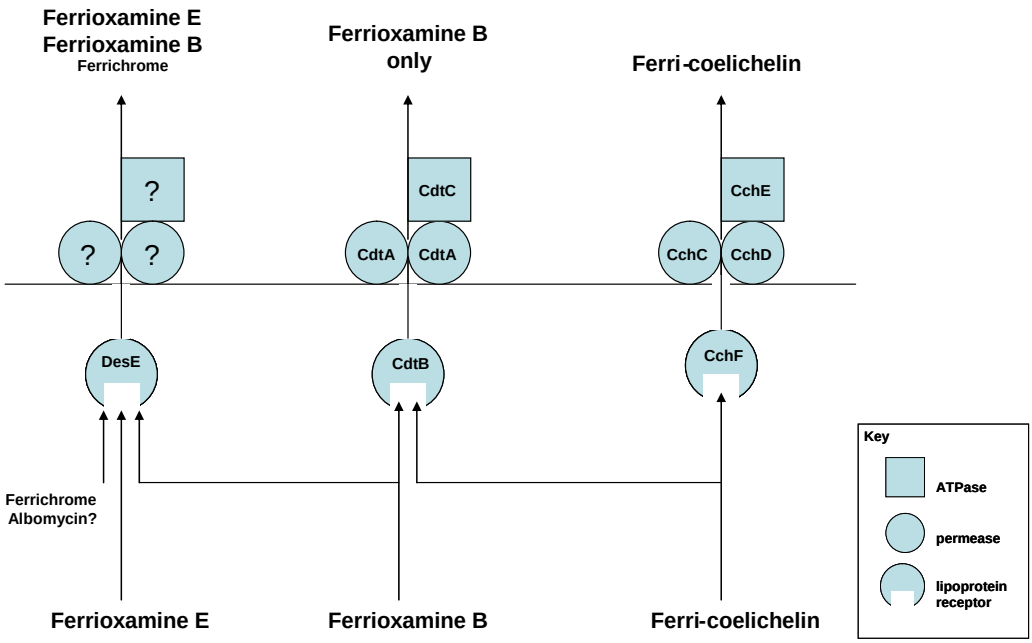


Figure 7-1: Proposed uptake pathways for ferri-siderophores in *S. coelicolor*.

Ferrioxamine E is transported into *S. coelicolor* cells via the lipoprotein receptor DesE, through an as yet unidentified permease component of an ABC transporter system. Most probably, ferrioxamine B can also use this route due to the ability of DesE to bind this siderophore, the structural similarity of ferrioxamine B to ferrioxamine E and the fact that *S. coelicolor* mutants lacking the *cdt* gene cluster appear to be able to uptake ferrioxamine B[59].

The CchF lipoprotein can only bind ferri-coelichelin. Hence the ABC transporter consisting of CchF, the permeases CchC and CchD, and ATPase CchE, can only transport ferri-coelichelin. This suggests that the *cch* cluster, which is also found in *S. ambofaciens*, encodes a complete siderophore biosynthetic-export-uptake pathway, that can uptake iron via coelichelin independently of other siderophore biosynthesis-export-uptake systems.

The specificity of CdtB for ferrioxamine B suggests the *cdt* cluster is involved in the uptake of this ferri-siderophore. CdtB had surprisingly relatively little affinity for the structurally very similar ferrioxamine E, demonstrating remarkable substrate discrimination by this protein. This may indicate different modes of binding of ferrioxamine B to CdtB and DesE. The role of the *cdt* gene cluster in ferrioxamine B uptake warrants further investigation by deletion of the *cdt* and *des* clusters in the same strain of *S. coelicolor*, and conducting uptake experiments with radiolabelled ferrioxamine B, as well a growth promotion experiments in iron-deficient media supplemented by desferrioxamine B.

The nanomolar dissociation constant of ferri-coelichelin-CdtB complex was surprising, because the *cchCDEF* genes were found to be necessary for growth of coelichelin non-producing mutants supplemented by ferri-coelichelin[53], and not the *cdt* cluster. Hence at least some of these genes were required for efficient ferri-coelichelin uptake. The fate of CdtB-ferri-coelichelin could be investigated by growth or uptake assays with ferri-coelichelin involving non-siderophore producing mutants of *S. coelicolor* with the *cdtB* and *cchCDE* genes present to see whether the permease-ATPase genes in the *cch* cluster can release and transport ferri-coelichelin from CdtB (Figure 7-2).

It is possible that actually CdtB does not bind to ferri-coelichelin at all, and the observed binding was due to the instability of the protein. For instance, the protein may unfold slightly and ferri-coelichelin would be able to bind to the modified binding pocket, whilst the stable protein in *S. coelicolor* is unable to unfold to accommodate ferri-coelichelin. This could be investigated by comparing the kinetics of the formation of the CdtB-ferri-coelichelin to CdtB-ferrioxamine B complexes to see whether there are several steps involved in ferri-coelichelin binding than for

ferrioxamine B binding, or by solving the structure of CdtB-ferri-coelichelin by X-ray crystallography.

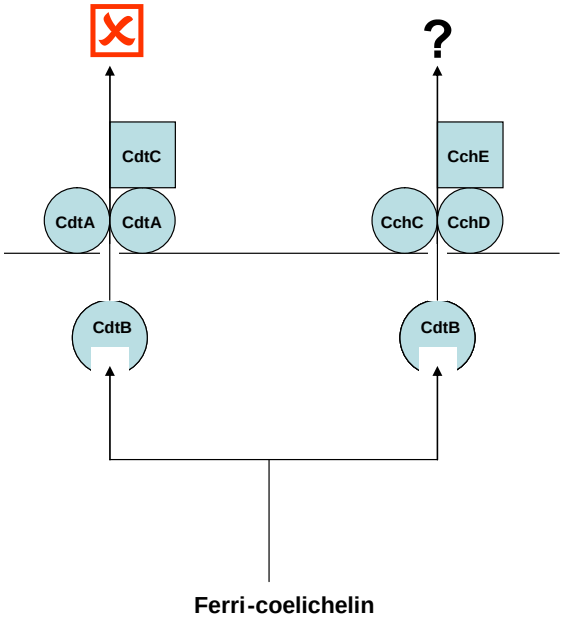


Figure 7-2: Possible fate of CdtB-ferri-coelichelin?

CdtB was found to bind ferri-coelichelin strongly, but the *cch* cluster was found to be necessary for growth of siderophore non-producing mutants supplemented with ferri-coelichelin. CdtB may substitute for CchF in uptake of this siderophore.

7.2.2 : Xenosiderophore uptake in *S. coelicolor*

The tight binding of the fungal siderophore ferrichrome to DesE and the evidence of Barona-Gómez *et al.* that the *des* cluster was necessary for efficient growth with this ferri-siderophore suggests the same transporter is used for uptake of ferrichrome and ferrioxamine E.

The antibiotic albomycin (a ferrichrome analogue) binds to the DesE protein with two orders of magnitude less affinity than ferrichrome. However, the bacterium is not sensitive to this antibiotic. This may indicate two things: (1) the binding of albomycin to DesE is not strong enough to allow efficient uptake, or (2) albomycin can be taken up but is deactivated or exported. To answer this question, one needs to determine whether albomycin can be taken up by *S. coelicolor* (uptake experiments with radiolabelled albomycin) or whether it can be used as an iron source in wild-type *S. coelicolor* and *des* cluster mutants.

The double *des* and *cch* cluster mutant of *S. coelicolor* was found to grow well when supplemented with the siderophore desferrioxamine G (an intermediate in the biosynthesis of ferrioxamine E in *S. coelicolor*[55]) as with ferric coprogen and the *E. coli* siderophore ferri-aerobactin. These are likely to be taken up by the *cdt* system as is ferrioxamine B (ferrioxamine G has a very similar structure), but it would be necessary to construct a siderophore non-producing mutant lacking the *des* and *cdt* clusters and carry out further growth promotion experiments. One could estimate dissociation constants for ferrioxamine G, ferric coprogen and ferri-aerobactin from DesE, CchF and CdtB to determine if these lipoproteins can bind these ferri-siderophores, but tight binding may not be indicative of siderophore transport as appears to be the case with CdtB-ferri-coelichelin.

Other putative lipoprotein receptors may be involved in xenosiderophore uptake in *S. coelicolor*; in particular SCO0996, SCO2272 and SCO7218 are strong candidates. Recombinant expression and purification of these proteins could be accomplished in a similar manner to those of DesE, CchF and CdtB. The main problem is that the roles of these proteins in iron metabolism in *S. coelicolor* is unknown, and they have not been observed in the membrane-bound proteome of *S. coelicolor*[62,63].

7.2.3 : Modelling Fe-siderophore – siderophore-binding-protein binding

The X-ray crystal structures of apo-DesE and the DesE-ferrioxamine B complex demonstrates that the conformation of DesE does not change on substrate binding but the configuration of the substrate does. This highlights an important consideration for modelling the binding of ferri-siderophores to siderophore-binding proteins of bacteria – the dynamics of ferri-siderophore complexes have to be modelled effectively.

In this study, a forcefield in the program DOMMIMOE[73] that handles ferric-*tris*-hydroxamates was constructed where trigonal distortion was sufficiently implemented. Hence the energy needed to affect configurational change of ferrioxamine B, for example, to allow binding to a siderophore-binding protein such as DesE could be calculated. This model could be used with the molecular dynamics

features of DOMMIMOE to calculate values such as dissociation constants and absolute binding energies, to assist in trying to understand the molecular basis of siderophore-SBP binding.

Chapter 8: Experimental methods

8.1 : Buffers and General solutions

8.1.1 : Chemicals and Reagents

All chemicals which were used in this study were of AnalR or molecular biology grade and were purchased from Sigma, Difco, Fisher or VWR unless otherwise stated.

Molecular biology reagents were used as in the manufacturers' guidelines. Distilled water was used for all solutions and media unless stated otherwise. Sterilization was achieved by autoclaving (121 °C, 20 min; Priorclave). Solutions unable to be autoclaved were passed through disposable 0.2 μm pore size cellulose acetate sterile filters (Millipore).

8.1.2 : Instruments and Equipment

Electroporations were performed using a BIORAD Genepulser II electroporator.

Centrifugation was performed using Eppendorf centrifuge 5415D, Avanti J-555. PCR reactions were performed on an Eppendorf Mastercycler personal thermocycler. Other instruments used: Priorclave autoclave, Gen lab incubator, New Brunswick Scientific C24KC refrigerated incubator shaker. Lyophilization was performed using a VirTis Benchtop K freeze-dryer.

ESI-MS was performed on a Bruker micrOTOF machine and LC-MS performed using an Agilent 1100 machine connected via a splitter to a Bruker Esquire HCTplus ESI-MS-MS. Preparative HPLC was performed on a quad-pump Agilent 1100 series machine run by Bruker HyStar software.

8.1.3 : Agarose Gel Electrophoresis buffers

2x TBE stock solution buffer was prepared using a standard protocol[80]. 10x gel loading buffer was prepared with glycerol (30% in H_2O), bromophenol blue (0.25%), xylene cyanol FF (0.25%) using a standard protocol[80].

Agarose electrophoresis gels between 1% and 1.25% (w/v) were prepared from electrophoresis grade LE agarose (SeaKem) and mixed with 1x TBE buffer. The solution was heated by microwave until all the agarose had dissolved and ethidium bromide (final concentration 2 μg/mL) was added and poured into the gel mould with the appropriate gel comb and left to cool.

8.1.4 : Protein purification buffers

The protein buffers were prepared using distilled water and filtered in 0.22 μm cellulose acetate filters before use and stored at 4 °C.

Washing buffer

10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 20 mM imidazole, 10% v/v glycerol

Elution buffer

10mM Tris-HCl (pH 8.0), 100 mM NaCl, 300 mM imidazole,10% v/v glycerol

Gel Filtration buffer

10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10% v/v glycerol

8.1.5 : Fluorescence buffers

All buffers for fluorescence measurements were prepared in HPLC grade water filtered in 0.2 μm pore cellulose acetate filters (Millipore) before use to remove small particles which may affect fluorescence.

Sodium Phosphate buffer (pH 7.0)

10 mM $Na_2(PO_4)_3$ buffered with 0.1 M NaOH and 25% H_3PO_4 (v/v) .

Phosphate Buffered Saline (PBS) (pH 7.0)

10 mM $Na_2(PO_4)_3$ buffered with 0.1 M NaOH and 25% H_3PO_4 (v/v) plus 100 mM NaCl.

8.1.6 : Chemical Stock Solutions

Stock solutions were prepared using the methods of Sambrook and Russell (Table 8-1)[80]

Table 8-1: Chemical stock solutions

Chemical	Stock solution	Solvent
Ethidium Bromide	10 mg/mL	water
10% glycerol	10% v/v	water
50% glycerol	50% v/v	water
NaOH	10 M	water
Ammonium persulfate	10% w/v	water
IPTG (iso-pyrano thio-β-galactosidase) (Melford)	1 mM	water

8.1.7 : Antibiotic stock solutions

All antibiotic stock solutions were prepared using distilled water and filter-sterilised by filtration (0.2 μm filters, Millipore) before use (Table 8-2).

Table 8-2: Antibiotic stock solutions

Antibiotic	Manufacturer	Stock solution	Working concentration
Ampicillin	Melford	50 mg/mL	100 μg/mL (for 5 mL overnight cultures and LB Agar) 50 μg/mL (for protein overproduction)
Kanamycin	DUCHEFA	50 mg/mL	100 μg/mL
Apramycin	DUCHEFA	50 mg/mL	100 μg/mL
Chlorophenamicol	DUCHEFA	30 mg/mL	30 μg/mL

8.1.8 : Microbial strains and vectors

8.1.8.1 : *Streptomyces* strains

S. coelicolor M145, lacking the SCP1 and SCP2 plasmids present in *S. coelicolor* A3(2), and *S. coelicolor* W2, a non-desferrioxamine B and desferrioxamine E producing mutant of *S. coelicolor* M145, were used to produce coelichelin (Table 8-3).

Table 8-3: *Streptomyces* strains used in this study

Strain	Genotype	Description
Streptomyces coelicolor M145 (wild-type)	*SCP1*$^-$ *SCP2*$^-$	*Streptomyces coelicolor A3(2)* strain without plasmids SCP1 and SCP2
Streptomyces coelicolor W2[53]	*SCP1*$^-$ *SCP2*$^-$ *desD::aac(3)IV*	*S. coelicolor* M145 non-desferrioxamine producing mutant. Apramycin resistant.
Streptomyces ambofaciens ATCC23877		
Streptomyces longisporus ruber		

8.1.8.2 : *Escherichia coli* strains

E. coli MC1061 (obtained from Professor G. Challis) and TOP10 (Invitrogen) were used for general storage of plasmids. *E. coli* BL21star (DE3) (Invitrogen) was used for the overproduction of proteins (Table 8-4).

Table 8-4: *E. coli* strains

Strain	Genotype	Description
Escherichia coli MC1061	*ara*D139 Δ(*araA-leu*)7697 Δ*lac*X74 *galK16 galE15 e14*$^-$ *mcrA0 relA1 rps*L *spo*T1 *mcr*B1 *hsd*R2 λ$^-$	Plasmid-holding strain
E. coli TOP10	F- *ara*D139 Δ(*ara-leu*)7697 *mcr*A Δ*lac*X74 *gal*K, Δ(*mrr-hsd*RMS-*mcr*BC) *deo*R *rec*A1 *rps*L *end*A1 *nup*G *gal*E15 *gal*K16 λ$^-$	Plasmid-holding strain supplied by Invitrogen in TOPO kit
E. coli BL21star (DE3)	F- *ompT hsdS*B (r_B $^-m_B^-$) *gal dcm rne131 λ*(DE3)	Overproducing strain supplied by Invitrogen in TOPO kit. Contains T7 polymerase.
E. coli	F *ompT hsdS* $^+$ (r_B	Tobacco etch virus (TEV)

BL21(DE3)-RIL / pRK793	$^{-}m_B^{-}$)Ter^r *gal* λ (DE3) *endA* Hte (*argU ileY leuW* Cam^r)	protease overproducing strain of *E. coli* transformed with plasmid pRK793. Kanamycin and chloramphenicol resistant.

8.1.8.3 : **Plasmids**

Table 8-5: Commercially available plasmids

Plasmid	Supplier	Host	Features
SCC105	John Innes Centre, Norwich, UK	*E. coli* MC1061	Contains *S. coelicolor* genes *sco2759 – sco2793*
SCF34	John Innes Centre, Norwich, UK	*E. coli* DH5α	Contains *S. coelicolor* genes *sco0482 – sco00505*

8.1.9 : Growth media and culture conditions

Inoculation of media with *Streptomyces sp.* and *Escherichia coli* were done under aseptic technique. For *E. coli* inoculation was performed under a Bunsen burner flame whilst for *Streptomyces* inoculation was performed under a flame in a laminar flow hood. All sterile solutions and media were prepared under laminar flow hood conditions.

Luria-Bertani (LB) Medium[80]

Tryptone (10g/L), yeast extract (5g/L) and NaCl (10g/L) or Luria-Bertani Broth powder (Lennox) (25g/L) was dissolved in distilled water and sterilized. 15g/L Bacto agar was added to make LB agar.

Soya Flour Mannitol Medium (SFM)[60]

Soya flour obtained from a health food shop (20g/L), D-mannitol (20g/L) and Bacto agar (15g/L) were dissolved in tap water and sterilized.

Iron Deficient Medium (for coelichelin production)[52]

Iron deficient medium was made on a recipe based on the medium of Müller and Raymond[12]. Salts were dissolved in distilled water (Table 8-6) and Chelex-100 resin (Sigma-Aldrich, 50g/L) was added and stirred overnight in plastic containers.

Table 8-6: Salt concentrations in iron-deficient medium

Ingredient	Concentration
K_2SO_4	2g/L
K_2HPO_4	3g/L
NaCl	1g/L
NH_4Cl	5g/L

The resin was filtered by gravity in a plastic funnel with Whatman No.1 filter paper and then trace nutrients were added to final concentrations given in Table 8-7 and autoclaved. To start the cultures, 10 mL per litre of sterile $CaCl_2$ (10g/L), sterile glucose (250g/L) and sterile yeast extract (0.5%) was added with the *Streptomyces* spores.

Table 8-7: Trace nutrient concentrations used in iron-deficient medium

Ingredient	Concentration
Thiamine	2 mg/L
$ZnSO_4.7H_2O$	2 mg/L
$CuSO_4$	5 µg/L
$MnSO_4.H_2O$	35 µg/L
$MgSO_4.H_2O$	80 mg/L

8.2 : Purification of ferri-coelichelin

8.2.1 : Preparation of spore stocks

Streptomyces coelicolor M145, *S. coelicolor* W2, *Streptomyces ambofaciens* ATCC23877 and *Streptomyces longisporus ruber* spore stocks were made by adding 5-10 µL of spore to 1 mL of sterile water on SFM plate (in the case of *S. coelicolor* W2 supplemented with 60 µg/mL apramycin) with incubating at 30 °C for 8-10 days until complete sporulation had occurred. The spores were harvested by applying 4-5

mL sterile water to the plate and scraping the hydrophobic surface to release spores into the water. The spore suspension was filtered through sterile glass wool to remove mycelial fragments. The filtrate was centrifuged (4000 rpm, 5 min, 4 °C) and the sediment was resuspended in sterile 50% glycerol to a final total concentration of 25% glycerol. The spore stocks were stored at -20 °C.

8.2.2 : Small-scale growth of *Streptomyces* in iron-deficient medium (50 mL)

50 μL of *Streptomyces* spores were added to 50 mL of iron-deficient medium and the culture was incubated in a rotary shaker (30 °C, 180 rpm) for 4-5 days.

8.2.3 : Large-scale growth of *Streptomyces coelicolor* in iron-deficient medium (up to 2 L)

500 μL of *Streptomyces* spores were added to 500 mL of iron-deficient medium and the culture was incubated in a rotary shaker (30 °C, 180 rpm) for 4-5 days. After this period, the culture developed a dark violet colour.

8.2.4 : Processing of culture supernatant

The supernatant was filtered through Whatman No.1 filter paper under partial vacuum and $FeCl_3$ was added to final concentration of 1 mM. The supernatant developed a reddish-violet colour and became cloudy. The sample was evaporated under vacuum (35 °C, 1-2 hours) until dry and the residue was dissolved in 5 mL H_2O. The concentrated supernatant was filtered through 0.2 μm cellulose HPLC filters (Millipore) prior to HPLC or LC-MS analysis.

8.2.5 : LC-MS analysis of culture supernatants[53]

An Agilent 1100 instrument equipped with a binary pump and a diode array detector was used for HPLC analyses. Samples were analysed on a Supelco Discovery HSF5 column (150 x 4.6 mm, 5 μm pore size, column temperature 20 °C) using the gradient solvent profile given in (Table 8-8):

The HPLC outflow was connected via a splitter (90% flow to waste; 10% flow to MS) to a Bruker HCT+ ESI mass spectrometer (Bruker Daltonics, Coventry). The spectrometer was operated in positive ion mode with parameters as follows: nebulizer flow 40 p.s.i., dry gas flow 10 L / min, dry temperature 300 °C, capillary –4 kV, skimmer 40 V, capillary exit 121 V, and ion charge control target (ICC) 100000, spectrum averages 3. The retention time of ferri-coelichelin was 3.2 minutes.

Table 8-8: HPLC eluent profile in LC-MS of culture supernatants

Time (min)	10 mM ammonium carbonate buffer (pH 7.0)	Methanol	Flow Rate (mL/min)
0	90%	10%	1
10	90%	10%	1
18	0%	100%	1
28	0%	100%	1
36	90%	10%	1
40	90%	10%	1

8.2.6 : RP-HPLC purification of ferri-coelichelin

The processed culture supernatant of large cultures of *S. coelicolor* M145 or W2 was separated by reversed-phase HPLC (Agilent Zorbax C-18 column, 21.2 x 10.0 mm, 5 μm particle size, temperature 20 °C) using an Agilent 1100 series instrument with a UV detector monitoring absorbance at 435 nm. The solvent elution profile is given in Table 8-9.

Table 8-9: HPLC eluent profile for purification of ferri-coelichelin

Time (min)	Water (0.1% TFA)	Acetonitrile (0.1% TFA)	Flow Rate (mL/min)
0	100%	0%	4.2
15	90%	10%	4.2
20	0%	100%	4.2
30	0%	100%	4.2
35	100%	0%	4.2
45	100%	0%	4.2

Fractions were collected and analysed to see if they contained ferri-coelichelin by direct injection into a Bruker micrOTOF ESI-MS operating in positive ion mode. The retention time of ferri-coelichelin was 7 - 9 minutes. The fractions containing ferri-coelichelin were snap-frozen in liquid N_2 and lyophilized.

8.2.7 : Removal of iron from ferri-coelichelin

Using the established protocol[75] partially purified ferri-coelichelin was dissolved in 25 mL of ultrapure water and 145 mg (1 mmole) of 8-hydroxyquinoline dissolved in 25 mL HPLC grade methanol and the two solutions were mixed together for 30 minutes, during which time the solution turned from light orange-yellow to green. All glassware used from this step onwards was rinsed in 6 M HNO_3 and deionised H_2O to remove traces of iron. The dark green ferri-8-hydroxyquinoline was removed by extraction with 20 mL dichloromethane (five times). The aqueous phase was retained, and excess organic solvents were evaporated under vacuum. The remaining aqueous solution was snap-frozen in liquid N_2 and lyophilized in iron-free glassware.

8.2.8 : Purification of desferri-coelichelin by RP-HPLC

Due to the instability of desferri-coelichelin, this step was carried out without delay. A published protocol was followed for this step[52]. After dissolving in 4 mL of 10 mM ammonium hydrogen carbonate buffer (pH 7.0) and filtering through 0.2 μm cellulose acetate filters, desferri-coelichelin was purified by reversed-phase HPLC (Phenomenex Synergi Fusion RP40 column, 250 x 10 mm, 4 μm particle size) using an Agilent 1100 series instrument with a UV detector monitoring absorbance at 210 nm. The solvent elution profile is given in Table 8-10.

Fractions were collected and analysed to see if they contained desferri-coelichelin by direct injection into a Bruker micrOTOF ESI-MS operating in positive ion mode. Fractions containing desferri-coelichelin were stored on ice, then combined and frozen in liquid N_2 and lyophilized.

Table 8-10: HPLC eluent profile for purification of desferri-coelichelin

Time (min)	10 mM ammonium hydrogen carbonate buffer (pH 7.0)	Acetonitrile	Flow Rate (mL/min)
0	100%	0%	3
25	100%	0%	3
30	0%	100%	3
40	0%	100%	3
45	100%	0%	3
55	100%	0%	3

8.2.9 : Addition of gallium to desferri-coelichelin

Using an established protocol[75], 151 mg $Ga_2(SO_4)_3$ was dissolved in 5 mL of 0.05 M H_2SO_4 and added to a 5 mL solution of desferri-coelichelin. After 30 min, the solution was neutralized with 0.5 M NaOH, frozen in liquid N_2 and lyophilized in plastic centrifuge tubes to prevent the loss of the light and fluffy gallium-coelichelin powder.

8.2.10 : Purification of gallium-coelichelin by RP-HPLC

The gallium-coelichelin was dissolved in 4 mL of 10 mM ammonium carbonate buffer (pH 7.0) and filtered through 0.2 μm cellulose filters. Gallium-coelichelin was purified to homogeneity by reversed-phase HPLC (Agilent Zorbax C-18 column, 21.2 x 10.0 mm, 5 μm particle size) using an Agilent 1100 series instrument with a UV detector monitoring absorbance at 210 nm. The solvent elution profile is given in Table 8-11.

Table 8-11: HPLC eluent profile for purification of gallium-coelichelin

Time (min)	10 mM ammonium hydrogen carbonate buffer (pH 7.0)	Acetonitrile	Flow Rate (mL/min)
0	100%	0%	5
15	100%	0%	5
20	0%	100%	5

30	0%	100%	5
35	100%	0%	5
45	100%	0%	5

Fractions were collected and analysed to see if they contained gallium-coelichelin by direct injection into a Bruker micrOTOF ESI-MS operating in positive ion mode. Fractions containing gallium-coelichelin (7-8 minutes retention time) were collected and lyophilized in 15 mL plastic centrifuge tubes to prevent the loss of the light and fluffy powder.

8.2.11 : ^{1}H NMR of gallium-coelichelin

Gallium-coelichelin was dissolved in 750 μL DMSO-d_6 (Cambridge Isotope Laboratories) and a ^{1}H NMR spectrum was measured at 400 MHz (Bruker Biospin, Coventry). Assignments of the signals observed are given in Chapter 2.

8.2.12 : COSY and ^{13}C -^{1}H HSQC of gallium-coelichelin

COSY and HSQC spectra were measured on a 700 MHz machine (Bruker Daltonics, Coventry) equipped with a TCI cryoprobe.

8.2.13 : Estimation of gallium-coelichelin concentration

60 μg CH_2Br_2 (Supelco) dissolved in 15 μL DMSO-d_6 was added to the NMR tube containing the gallium-coelichelin. Comparison of the integrals for the signal due to CH_2Br_2 (5.35 ppm) an the signal due to the proton on the α carbon of the L-threonine residue of gallium-coelichelin (4.38 ppm) gave the molar ratio of dibromomethane to gallium-coelichelin, allowing the concentration of gallium-coelichelin to be calculated from the known amount of CH_2Br_2 added.

8.2.14 : Conversion of gallium-coelichelin to ferri-coelichelin

1200 μL of a 1 mM $FeCl_3$ solution was added to a gallium-coelichelin solution and left to react for 15 h at 4 °C. The conversion of the gallium complex (m/z 632) to

the ferric complex (m/z 619) was monitored by direct injection into a Bruker micrOTOF ESI mass-spectrometer in positive ion mode.

8.2.15 : LC-MS analysis of pure ferri-coelichelin

The method used was as described previously[53], with 10 μL of a 100 μM solution of ferri-coelichelin injected onto the HPLC column.

8.2.16 : ESI-MS-MS analysis of ferri-coelichelin in positive ion mode

A dilute sample of ferri-coelichelin was injected into a Bruker HCT+ ESI mass spectrometer (Bruker Daltonics, Coventry) operating in positive ion mode with parameters as follows: nebulizer flow 40 p.s.i., dry gas flow 10 L / min, dry temperature 300 °C, capillary – 4 kV, skimmer 40 V, capillary exit 121 V, ion charge control target (ICC) 100000, spectra averages 3.

8.3 : Overproduction, isolation and characterization of recombinant proteins DesE, CchF and CdtB

8.3.1 : Miniprep extraction of plasmids/cosmids from *E. coli*

The standard miniprep procedure as described by Sambrook and Russell was used to isolate plasmid and cosmid DNA[80]. Alkaline Lysis solutions I, II and III were made as described by Sambrook and Russell[80]. To obtain DNA of sufficient quality for sequencing reactions, the QIAGEN miniprep kit was used. In both miniprep procedures, the isolated DNA was resuspended in EB buffer (10 mM Tris-Cl, pH 8.5).

8.3.2 : Preparation of primers

Primers were ordered from Sigma Genosys, dissolved in sterile water to a final concentration of 1 μg/mL and stored at -20 °C. The sequences of the primers used in this study are listed in Table 8-12.

8.3.3 : PCR (polymerase chain reaction) amplification of genes.

The template DNA used to amplify the gene *desE* was the *S. coelicolor* cosmid SCC103 (from Prof. G. L. Challis). The *S. coelicolor* cosmid SCF34 (from Prof. G. L. Challis) was used to amplify the gene *cchF*, and genomic *S. coelicolor M145* DNA (from Dr. C. Corre) was used to amplify the *cdtB* gene. Minipreps of plasmids pPP001 (pET151/D-TOPO + *desE*), pPP002 (pET151/D-TOPO + *cchF*) and pPP003 (pET151/D-TOPO + *cdtB*) were used to amplify the genes 5'-truncated *desE*, 5'-truncated *cchF* and 5'-truncated *cdtB* respectively.

Table 8-12: PCR primers used in this study. Bases in bold were required to facilitate TOPO cloning (CACC) or mutate Cys to Met (ATG).

Primer	Sequence
desE forward	**CACC**ATGTCCCACGCCAGC
truncated_desE forward	**CACCATG**GGCGACGGCGACGGCAAG
desE reverse	TCAGCCGACCTTCTTGGCGT
cchF forward	**CACC**ATGCTCCTCAGAACAA
truncated_cchF forward	**CACCATG**GGATCCGACTCGGACGA
cchF reverse	CTACCCGGCCGACTTGACGA
cdtB forward	**CACC**ATGAGACGCCTCCTGC
truncated_cdtB forward	**CACCATG**GGCACCACCGAACCCG
cdtB reverse	CTACTTCGTCAGCGCGCCGACGAC
T7 forward	TAATACGACTCACTATAGGG
T7 reverse	TAGTTATTGCTCAGCGGTGG

The primers *desE* forward and *desE* reverse were used to amplify *desE*; the primers *cchF* forward and *cchF* reverse were used to amplify *cchF*; the primers *cdtB* forward and *cdtB* reverse were used to amplify *cdtB*. The primers *truncated_desE* forward and *desE* reverse were used to amplify 5'-truncated *desE*; the primers *truncated_cchF* forward and *cchF* reverse were used to amplify 5'-truncated *cchF*; the primers *truncated_cdtB* forward and *cdtB* reverse were used to amplify 5'-truncated *cdtB*. PCR was performed using a high fidelity polymerase was used with the addition of 5% dimethyl sulfoxide (DMSO) due to the high GC content of the template DNA. PCR reactions contained the reagents listed in Table 8-13. The PCR

reactions were performed in an Eppendorf Mastercycler at annealing and elongation temperatures of 55 °C and 72 °C respectively, using the program given in Table 8-14.

Table 8-13: Reagents in PCR reactions

Reagent	µL
Sterile water	37.5
DMSO	2.5
Template DNA (~10 ng/µL)	1
dNTP mix (dATP, dTTP, dGTP, dCTP) (5 mM)	1
Forward primer (50 pmol)	1
Reverse primer (50 pmol)	1
10x PCR buffer (+ $MgCl_2$)	5
Expand High Fidelity polymerase (Roche)	1

Table 8-14: PCR program used for all PCR reactions

Step number	Temperature	Duration of step (seconds)
1	94 °C	120
2	94 °C	45
3	55 °C	45
4	72 °C	90
5	GOTO 2 REPEAT x30	
6	72 °C	900
7	4C	store

8.3.4 : Analysis and purification of PCR products

6 µL of PCR products were mixed with 1 µL gel loading buffer (6x MassRuler loading dye, Fermentas) and analysed on a 1 – 1.2% agarose gel by electrophoresis using 6 µL of GeneRuler™ 1 kb DNA ladder (Fermentas) as a molecular size standard. The gel was visualized on a gel documentation system (UV transilluminator, UVP) after ethidium bromide staining, and the concentration of the

desired amplimer was estimated by comparing the fluorescence intensity against bands in the DNA ladder. A larger sample of the PCR products (20 μL) was separated on an agarose gel and the DNA of the appropriate size was excised to separate the product from the template and unreacted nucleotides. The QIAGEN gel extraction kit was used to extract the DNA from the excised agarose plug, eluting with the buffer provided.

8.3.5 : Insertion of amplimers into pET151-D/TOPO

The manufacturers' protocol was used as described in the booklet provided with the kit "Champion™ pET Directional TOPO® Expression Kits": The reagents for the TOPO® cloning reaction are given in Table 8-15.

Table 8-15: Reagents used in the ligation of the pET151/D-TOPO plasmid to the amplimers

Reagent	Volume
Purified PCR product (approx. amount)	1 μL – 3 μL
Salt solution (kit)	1 μL
TOPO® vector (pET151-D/TOPO)	1 μL
Sterile water	to a final volume of 6 μL

The volume of purified PCR product used in the reaction depended on the estimated concentration of the product. The total amount of insert DNA used in the TOPO reaction was around 1 – 5 ng. The reaction was carried out at room temperature in a 200 μL sterile centrifuge tube and 3 μL of product was used to transform *E. coli* MC1061 and TOP10 competent cells.

8.3.6 : Transformation of *E. coli* BL21star and MC1061 strains with plasmids by electroporation

5 mL of LB medium was inoculated with *E. coli* BL21star or MC1061 cells, and the culture was grown overnight in a rotary shaker (37 °C, 180 rpm). 50 mL of LB medium was inoculated with 250 μL of the overnight culture, and the culture was

incubated at 37 °C and 180 rpm until it grew to OD_{600} ~ 0.4. The culture was centrifuged (10 min, 4 °C, 2700 rpm). The supernatant was decanted and the pellet was resuspended gently in 25 mL of 10% ice-cold glycerol. The cells were centrifuged again (10 min, 4 °C, 2700 rpm), the supernatant was decanted and the pellet resuspended again in 25 mL 10% ice-cold glycerol. The cells were centrifuged again (10 min, 4 °C, 2700 rpm), the supernatant was decanted and the pellet was resuspended in 500 µL of 10% ice-cold glycerol solution.

50 µL of cell suspension was mixed with 1 µL of plasmid DNA or TOPO® reaction DNA in an ice-cold electroporation cuvette. Electroporation was carried out using a BioRad Gene Pulser II instrument set to 200 Ω, 25 uF and 1.9 kV. 1 mL of ice-cold LB was immediately added to the cells which were incubated (45 min, 180 rpm, 37 °C) before spreading onto LB agar containing 100 µg/mL ampicillin and incubating overnight at 37 °C.

8.3.7 : Transformation of *E. coli* One-Shot BL21star and TOP10 chemically competent cells

50 µL aliquots of chemically competent *E. coli* cells (Invitrogen) stored at -80 °C were thawed on ice, and mixed with 2 µL plasmid DNA (or 3 µL TOPO® reaction) and incubated on ice for 25-30 minutes. The cells were heat shocked at 42 °C for 30 seconds, incubated on ice for 2 minutes and resuspended in 200 µL of LB medium. This suspension was incubated (45 min, 180 rpm, 37 °C) before being spread onto LB agar supplemented with 100 µg/mL ampicillin and incubated overnight at 37 °C.

8.3.8 : PCR analysis of cloning reactions

Single colonies of *E. coli* were inoculated into 5 mL LB (100 µg/mL ampicillin) and grown overnight (37 °C, 180 rpm). The plasmids from the cultures were extracted by the standard miniprep protocol and the PCR reaction was performed on the plasmid DNA with the original primers used for coding sequence amplification using *Taq* polymerase. The PCR reactions were analysed by 1.2% agarose gel electrophoresis and clones containing the correct inserts (about 1 kbps) were selected for analysis of protein overproduction.

8.3.9 : Expression analysis of clones

Plasmid DNA was used to transform *E. coli* BL21star. A single colony was picked and grown overnight (5 mL LB, 50 μg/mL ampicillin, 37 °C, 180 rpm) and 250 μL of the culture was used to inoculate 50 mL LB (50 μg/mL ampicillin). The culture was grown for approximately 3-4 hours (37 °C, 180 rpm) until it reached OD_{600} ~ 0.6 and then it was stored on ice. A sterile solution of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration 0.5 mM for protein overproduction, and the culture was incubated in a rotary shaker overnight (15 °C, 180 rpm).

8.3.10 : Preparation of total protein fraction

1 mL of the 50 mL culture was centrifuged (14,000 rpm, 10 min) and the supernatant decanted. The pellet was resuspended in 100 μL cell lysis buffer[80] and boiled for 100 °C for 10 minutes.

8.3.11 : Preparation of soluble protein fraction (cell-free extract)

35 mL of overproduction culture was centrifuged (20 min, 14,000 rpm, 4 °C) and resuspended in 4 mL of washing buffer and 40 μL of a 1 mM phenylmethylsulfonylfluoride (PMSF) solution was added to inhibit proteases. The cells were lysed by pulsed sonication on ice (Sonics Vibracell, 2 min, 40 dB, 1 sec pulse on, 1 sec pulse off) and centrifuged (20 min, 14,000 rpm, 4 °C) to pellet cell debris and unlysed cells. The cell-free extract was decanted and stored at 4 °C.

8.3.12 : Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis of protein fractions

The total and soluble protein fractions were mixed with 5x SDS-PAGE loading buffer [80] and were analysed on an 8% SDS-PAGE gel using a Mini-Protean II gel electrophoresis system (BioRad). Either a bovine serum albumin (BSA)/ovalbumin (OVA) marker or PageRuler™ Plus Prestained Protein ladder (Fermentas) was used as molecular weight standards. The gel was stained using Coomassie Blue[80] and

destained by heating in the microwave for 10 min at 700 W in distilled water. The gel was visualized on a gel documentation system (UV transilluminator, UVP).

8.3.13 : Sequencing of the gene inserts of plasmids

Plasmid DNA prepared from holding strains of *E. coli* using the standard procedure[80]. The forward and reverse primers used were T7 forward primer and T7 reverse primer, which flank the inserted gene region. The DNA sequencing was carried out by the Biological Sciences Sequencing Service (University of Warwick) and GATC Biotech (Germany) and chromatograms analysed by Chromas 1.45 for mutations.

8.3.14 : Large-scale protein overproduction

2 mL of an overnight *E. coli* BL21star culture (containing the expression clone of interest) was inoculated in 600 mL LB (50 μg/mL ampicillin). The culture was grown at 37 °C for 3-4 hours until and OD_{600} ~ 0.6 was reached. IPTG was added to a final concentration of 0.5 mM and the culture was incubated in a rotary shaker overnight (15 °C, 180 rpm).

The overnight culture was centrifuged (20 min, 10000 rpm, 4 °C) and the pellet was recovered and stored at -80 °C overnight (in the case of protein CdtB to facilitate lysis). The pellet was resuspended in 20 mL washing buffer containing 10 μM PMSF and the cells were lysed using a French Pressure Cell Press (Thermo Scientific) at 4 °C using the 40K Standard Cell (Thermo Scientific) and a pressure of 20,000 psi.

The cell-free extract was separated from the cell debris by centrifugation (20 min, 10,000 rpm, 4 °C) and stored on ice.

8.3.15 : Purification of proteins by Immobilized Metal Affinity Chromatography (IMAC)

Ni-IMAC of the cell free extract was performed on an ÄKTA purifier (Amersham Biosciences, GE Healthcare). The cell-free extract was injected via a 50 mL superloop (Pharmacia Biotech) at 0.35 mL/min onto a 1 mL HisTrap Ni-NTA sepharose column (GE Healthcare) at 4 °C. Non-specifically binding proteins were washed off the column with 15 mL 98% washing buffer/2% elution buffer. The His-tagged recombinant protein was eluted with 50 – 100 % elution buffer at 0.35 mL/min and collected in 5 mL glass tubes. Absorbance was monitored at 280 nm.

Combined fractions containing the recombinant protein were filtered through a 30,000 MW cutoff 15 mL Amicon Ultra centrifugal filter (Millipore), diluting with gel filtration buffer by centrifuging for approximately 25 minutes (3000 rpm, 4 °C). The purity of the protein was assessed by SDS-PAGE (8% gel) and the protein was aliquoted into 200 μL clean thin-walled centrifuge tubes and stored at -80 °C.

8.3.16 : Bradford assay for protein concentration determination

The Bradford reagent required for the assay was prepared according to a standard procedure[80]. Dilutions of the protein were prepared (0, 10%, 20%, 40%, 60% 80%, 90%, 100%) in parallel with dilutions of a 0.78 mg/mL BSA solution (determined by absorbance using the extinction coefficient of $\varepsilon_{280} = 0.667$ $(mg/mL)^{-1}cm^{-1}$) in a total volume of 100 μL. 1 mL of Bradford reagent was added to the protein solutions, which caused a colour change from brown to peacock blue on increasing concentration of protein. The colour change was monitored by absorbance at 595 nm in a 1 cm path length cuvette using a BioMate UV-Vis spectrophotometer (Thermo Scientific). The concentrations of the protein solutions were determined by comparison with the absorbances of the standard BSA curve (plotted in Microsoft Excel).

8.3.17 : Purification of tobacco etch virus (TEV) protease

E. coli BL21(DE3)-RIL transformed with pRK793 overproduced a autoinactivation-resistant S219V mutant recombinant TEV protease catalytic domain with an N-

terminal His-tag and a C-terminal polyarginine tag, and was the (http://mcl1.nclfcrf/gov/waugh.html). This strain was inoculated into 5 mL LB (100 μg/mL ampicillin, 30 μg/mL chloramphenicol) and incubated overnight at 37 °C.

2 mL of overnight culture was inoculated into 600 mL of LB (50 μg/mL ampicillin, 15 μg/mL chloramphenicol) and incubated in a rotary shaker (3 h, 37 °C, 180 rpm) until an OD_{600} ~ 0.6 was reached. IPTG was added to a final concentration of 0.5 mM and incubated overnight (3 h, 37 °C, 180 rpm). The protein was extracted and purified using the protocol described in sections 8.3.14 : and 8.3.15 :, and 50 μL aliquots stored at -80 °C.

8.3.18 : Digestion of proteins with TEV protease

Approximately 50 μL of TEV protease was added to 300 – 400 μL of purified His-tagged protein (approximately 1:50 ratio TEV protease: His-tagged protein) and the mixture was incubated for 12 -15 h at 4 °C.

8.3.19 : Purification of TEV protease digested protein

0.5 – 1 mL of digestion product was injected via a 2 mL injection loop onto a 1 mL HisTrap column at 0.15 mL/min with washing buffer as the mobile phase and the flow-through was collected and concentrated. Undigested protein and TEV protease were assumed to bind to the column. The undigested and digested products were analysed by SDS-PAGE.

8.3.20 : Peptide mass fingerprinting of proteins

Electrospray ionization mass spectrometry (ESI-MS) of tryptic digests of proteins excised from SDS-PAGE gels was performed by The Biological Mass Spectrometry and Proteomics Facility in the Department of Biological Sciences, University of Warwick.

8.3.21 : ESI-MS analysis of proteins

Proteins were dialysed five times in 10 mM ammonium acetate (pH 7.0) using a 10,000 MW spin filter (Vivaspin, Millipore Inc.) to give a final concentration of approximately 25 μM. The concentration of the protein solutions were determined by the Bradford assay. The solution was acidified by a small amount of 50% MeCN (0.1% TFA). A total volume of 60 μL of protein solution was injected continuously a Bruker micrOTOF ESI-MS (Bruker Daltonics, Coventry, UK) using the following settings: positive ion mode, scan range m/z 800– 5000, capillary exit voltage 250.0 V, skimmer voltage 90.1 V, hexapole 24.8 V, hexapole RF voltage 750.0 V, corrector fill 48 V, pulsar pull 369 V, pulsar push 369 V, reflector 1300 V, flight tube 9000 V, detector TOF 2140 V. The spectra were analysed and deconvoluted to identify multiply-charged species using Bruker Daltonics DataAnalysis 3.3 software (Bruker Daltonics, Coventry, UK).

8.3.22 : Gel filtration (size exclusion chromatography) of proteins

500 μL of a solution of protein was loaded onto a size exclusion column (HiLoad™ 75 Superdex™ 12 prep grade serin, column volume 110 mL, Amersham Biosciences) equilibrated with gel filtration buffer at 0.8 mL/min on an ÄKTA purifier (Amersham Biosciences, GE Healthcare). The column was eluted with isocratic gel filtration buffer. 5 mL fractions were collected in the case of preparative gel filtration. Absorbance was monitored at 280 nm. The column was calibrated with the Kit for Molecular Weights 12,000-200,000 Da (Sigma), consisting of Cytochrome c (12,400 Da), Carbonic Anhydrase (29,000 Da), Bovine Serum Albumin (66,000 Da), Alcohol Dehydrogenase (150,000 Da), β-Amylase (200,000 Da), and Blue dextran (2,000,000 Da), which eluted at volumes 89.2 mL, 71.5 mL, 59.6 mL, 56.1 mL, 52.9 mL and 40.1 mL, respectively.

8.4 : Binding of siderophores to siderophore binding proteins

8.4.1 : Preparation of ferri- and desferri-siderophore solutions (except ferri-coelichelin and ferrioxamine B)

The ferri-siderophores ferrioxamine E, ferrichrome, ferri-enterobactin and ferri-albomycin (mixture of δ_1, δ_2 and ε forms) and the desferri-siderophores desferrioxamine E, desferri-ferrichrome and enterobactin were purchased in 1 mg amounts from EMC Microcollections (Tübingen, Germany). 1 mg of each compound was dissolved in 100 µL ultrapure water and stored at -20 °C.

8.4.2 : Preparation of ferrioxamine B solution

56 mg deferoxamine B mesylate (Sigma) was dissolved in distilled water and an excess of 100 mM $FeCl_3$ was added. The solution changed colour from colourless to bright red due to the formation of ferrioxamine B. The water was removed by evaporation under a partial vacuum, and the residue was dissolved in 1 mL HPLC grade H_2O and filtered.

The conditions for purification of ferrioxamine B by HPLC were the same as previously described for ferri-coelichelin (see section 8.2.6 :). An orange fraction was collected and was verified to contain ferrioxamine B by ESI-MS in positive ion mode (m/z 614; ferrioxamine B + H^+). The fraction was evaporated to remove organic solvents, snap-frozen in liquid N_2 and lyophilized. The resulting dark red powder was redissolved in 1 mL of ultrapure water and stored at -20 °C.

8.4.3 : Concentration estimation of ferri-siderophores by UV-Visible spectroscopy

The concentration of ferri-siderophore solution was estimated by UV-Visible spectrometry in a 1 mm path length quartz cuvette (Starna Scientific UK) using a Jasco V550 spectrophotometer. The extinction coefficients used are displayed in Table 8-16.

Table 8-16: Ferri-siderophore extinction coefficients.

Siderophore	Wavelength (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Reference
Ferrioxamine B	428	2800	[11]
Ferrioxamine E	430	2750	[11]
Ferrichrome	425	2895	[11]
Albomycin	425	2895	EMC Microcollections
Ferri-enterobactin	495	5600	[11]

8.4.4 : Inductively coupled plasma optical emission spectroscopy (ICP-OES) of ferri-coelichelin and other ferri-siderophores

A Perkin Elmer Optima 5300 DV Optical Emission Spectrometer was used in analyses to calculate the concentration of Fe in 1/1000 dilutions of siderophore solutions. 40 μL of stock solutions of ferrioxamine E, ferrioxamine B, albomycin and ferrichrome of approximately 500 μM and 100 μM and ferri-coelichelin of approximate concentration 300 μM were diluted in 3.96 mL of ultrapure water prior to analysis. The Fe standards were prepared from dilutions of an Iron for AAS standard solution (1001 ± 3 mg/L) in ultrapure water and prepared in 15 mL plastic centrifuge tubes. The iron standards prepared were 0 ppm, 0.05 ppm, 0.1 ppm, 0.5 ppm, 1 ppm, 5 ppm, 10 ppm and 50 ppm. The emission wavelengths analysed were 238.204, 239.562 and 259.939 nm and the plasma temperature was 8000 °C. The emission intensities recorded were entered into Microsoft Excel and its linear regression function was used to analyse the relationship between the concentrations of standard solutions against emission intensity, with the intercept of the line set at zero. The relationship formulae were used to determine the concentration of Fe in the siderophore solutions and hence the concentration of siderophore (as the Fe-siderophore complex is one-to-one) in the stock solutions from their emission intensities.

8.4.5 : Fluorescence measurements

Proteins were dissolved in PBS to a final concentration of 0.78 mg/mL. In the case of CchF it was incubated at 37 °C for 45 minutes prior to dissolving in PBS.

Fluorescence of the proteins was measured using a Perkin Elmer LS 50B luminescence spectrometer with an excitation wavelength of 280 nm. The xenon flash lamp was the source of radiation, which illuminated at a frequency of 50 Hz. No cutoff filters were used in the course of the experiments. The fluorescence spectrum was recorded from 310 nm to 380 nm, with a scan speed of 50 nm/min. The data was recorded in increments of 0.5 nm, with each data point being an average of 12 measurements. The excitation and emission slit lengths were 5.0 nm and 8.0 nm respectively, except in the case of CchF where they were 5.0 nm and 10.0 nm respectively. The measurements were taken using a 3 mm path length quartz cuvette (Starna Scientific) with the total volume ranging from 90 – 100 μL. Siderophore solutions were titrated into the protein solutions, mixed by agitation of a micropipette tip and left to incubate in the cuvette for 4 minutes at 21±1 °C before recording the spectra. The titrations were repeated at least three times.

8.4.6 : Processing of fluorescence data

The fluorescence data was transferred into Microsoft Excel from their format, and collated into one worksheet. The fluorescence at 335 nm was used to monitor the relative proportion of binding of different concentrations of ferri-siderophore ligand to DesE and CdtB. In the case of CchF, the fluorescence was monitored at 325 nm.

The data was linearly transformed into proportion bound by the relation $\frac{F_0 - F}{F_0 - F_{\min}}$ in Microsoft Excel, where F_0 is the fluorescence of unbound protein, F is the fluorescence of the current protein-ligand solution and F_{min} is the fluorescence of the bound protein. The transformed data was copied into Origin 6.0 Professional[86] and the nonlinear curve fitting function was used to fit the data using the function Onesitebind2 with the free parameter K_D and fixed parameter $[P_T]$ the protein concentration in nM.

If F_{min} had to be estimated, as in the case of ferri-siderophore binding to CchF and CdtB, a value of F_{min} was used provisionally as CchF bound to high concentrations of ferri-coelichelin and CdtB bound to high concentrations of ferrioxamine B. The data was linearly transformed into “relative fluorescence decrease compared to

F_{min}(siderophore)" by the relation $\frac{F_0 - F}{F_0 - F_{min}}$ in Microsoft Excel. The transformed data was copied into Origin 6.0 Professional[86] and the nonlinear curve fitting function was used to fit the data using the function Onesitebind_lipid with the free parameters K_D and $Q = \frac{F_0 - F}{F_0 - F_{min}(siderophore)}$, and the fixed parameter $[P_T]$ (the protein concentration in nM).

Origin 6.0 Professional was also used to graph the relative binding curves of ferri-siderophores to DesE, CchF and CdtB (relative binding versus total ligand concentration in μM), with each data point being the mean of at least three separate measurements, with y-error bars indicating ±1 standard deviation. A curve of best fit was also drawn, which represents the curves of best fit of K_D (and Q) to the models Onesitebind2 and Onesitebind_lipid.

8.4.7 : Circular dichroism spectroscopy of proteins with siderophores

Proteins were dissolved in sodium phosphate buffer (pH 7.0) to a final concentration 80 μg/mL three hours prior to analysis. Ferri-siderophores were added to a final concentration of 5 μM and incubated for 5 minutes at 21 °C prior to analysis. The instrument used was a Jasco J715 spectropolarimeter with excitation slit length of 1 nm. A 1 mm path length quartz cuvette was used to hold the sample. The spectra were recorded from 190 – 250 nm and eight scans were accumulated, with the spectrum of the solvent subtracted upon processing.

8.5 : Molecular modelling of ferric hydroxamates

8.5.1 : Hardware

All DFT calculations were performed on a dual Intel Pentium Xeon workstation 3.0 GHz running Mandriva Linux using the Centre for Scientific Computing delivered desktop (University of Warwick). All other calculations were performed on a Viglen Intel Pentium 4 workstation 3.0 GHz running Mandriva Linux using the Centre for Scientific Computing delivered desktop (University of Warwick).

8.5.2 : DFT

DFT calculations were carried out using Amsterdam Density Functional 2005 program[114]. All calculations were run spin unrestricted, using triple zeta basis sets. The 2p core was frozen on Fe and the 1s orbital frozen on C, N and O. All calculations used COSMO (conductor-like screening model) solvation corresponding to water (probe radius 1.4 Å, dielectric constant 78.4), with solvation radii of Fe, O, N, C and H being 2.20, 1.30, 1.40, 2.40 and 1.16 respectively.

Most of the ferric tris-hydroxamate geometry optimizations and single point calculations were performed with no symmetry except for trigonal prismatic calculations which were performed with symmetry C_{3v}. All geometry optimizations were carried out using the generalized gradient approximation Perdew-Wang 91[115] (PW91) with default geometry and SCF convergence criteria.

8.5.3 : Molecular Mechanics

Molecular mechanics was performed using the DOMMIMOE program[73] implemented in the Molecular Operating Environment (MOE)[74], which uses traditional molecular mechanics for atoms not based around the metal centre, and a computational treatment of the angular overlap model (AOM) on the transition metal centre and the immediate coordination environment.

The MMFF94 forcefield[98] was used for calculations, with additional terms added to remove angles, torsions and stretch-bends and adjust the charge at the Fe(III) metal centre. Out-of-plane terms were added to keep the hydroxamate ligands planar. All other forcefield terms were left unchanged.

Parameter optimization and molecule modification were implemented in programs in scientific vector language (SVL), MOE's internal programming language.

REFERENCES

(1) Drechsel, H.; Winkelmann, G. *Iron Chelation and Siderophores: Transition Metals and Microbial Metabolism*; Harwood Academic Publishers, 1997.

(2) Weinberg, E. D. *Perspect. Biol. Med.* **1997**, *40*, 578-583.

(3) Krewulak, K. D.; Vogel, H. J. *Biochem. Biophys. Acta.* **2008**, *1778*, 1781-1804.

(4) Raymond, K. N.; Dertz, E. A.; Kim, S. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3584-3588.

(5) Gray-Owen, S. D.; Schryvers, A. B. *Trends Microbiol.* **1996**, *4*, 185-191.

(6) Zhu, H. Z.; Alexeev, D.; Hunter, D. J. B.; Campopiano, D. J.; Sadler, P. J. *Biochem. J.* **2003**, *376*, 35-41.

(7) Tong, Y.; Guo, M. *Arch. Biochem. Biophys.* **2008**, *481*, 1-15.

(8) Pilpa, R. M.; Fadeev, E. A.; Villareal, V. A.; Wong, M. L.; Phillips, M.; Clubb, R. T. *J. Mol. Biol.* **2006**, *360*, 435-447.

(9) Miethke, M.; Marahiel, M. A. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413-451.

(10) Wandersman, C.; Delepelaire, P. *Annu. Rev. Microbiol.* **2004**, *58*, 611-647.

(11) Matzanke, B. F.; Müller-Matzanke, G.; Raymond, K. N. *Iron Carriers and Iron Proteins: Siderophore-Mediated Iron Transport*; VCH, 1989.

(12) Müller, G.; Raymond, K. N. *J. Bacteriol.* **1984**, *160*, 304-312.

(13) Crisponi, G.; Remelli, M. *Coord. Chem. Rev.* **2008**, *252*, 1225-1240.

(14) Cane, D. E.; Walsh, C. T.; Khosla, C. *Science* **1998**, *282*, 63-68.

(15) Challis, G. L. *ChemBioChem* **2005**, 601-611.

(16) Winkelmann, G.; Drechsel, H. *Biotechnology (2nd edition), Chapter 5: Microbial Siderophores*, 1999.

(17) Braun, V. *Front. Biosci.* **2003**, *8*, S1409-S1421.

(18) Challis, G. L.; Hopwood, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *25*, 14555-14561.

(19) Fischbach, M. A.; Lin, H. N.; Liu, D. R.; Walsh, C. T. *Nature Chem. Biol.* **2006**, *2*, 132-138.

(20) Flores, F. J.; Martin, J. F. *Biochem. J.* **2004**, *380*, 497–503.

(21) Farhana, A.; Kumar, S.; Rathore, S. S.; Ghosh, P. C.; Ehtesham, N. Z.; Tyagi, A. K.; Hasnain, S. E. *PLoS One* **2008**, *3*, 2087.

(22) Buchanan, S. K.; Smith, B. S.; Venkatramani, L.; Xia, D.; Esser, L.; Palnitkar, M.; Chakraborty, R.; van der Helm, D.; Deisenhofer, J. *Nature Struct. Biol.* **1999**, *6*, 56-63.

(23) Ferguson, A. D.; Braun, V.; Fiedler, H. P.; Coulton, J. W.; Diederichs, K.; Welte, W. *Protein Science* **2000**, *9*, 956-963.

(24) Carter, D. M.; Miousse, I. R.; Gagnon, J. N.; Martinez, E.; Clements, A.; Lee, J.; Hancock, M. A.; Gagnon, H.; Pawelek, P. D.; Coulton, J. W. *J. Biol. Chem.* **2006**, *281*, 35413-35424.

(25) Davidson, A. L.; Dassa, E.; Orelle, C.; Chen, J. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 317-364.

(26) Hvorup, R. N.; Goetz, B. A.; Niederer, M.; Hollenstein, K.; Perozo, E.; Locher, K. P. *Science* **2007**, *317*, 1387-1390.

(27) Karpowich, N. K.; Huang, H. H.; Smith, P. C.; Hunt, J. F. *J. Biol. Chem.* **2003**, *278*, 8429-8434.

(28) Clarke, T. E.; Braun, V.; Winkelmann, G.; Tari, L. W.; Vogel, H. J. *J. Biol. Chem.* **2002**, *277*, 13966-13972.

(29) Spurlino, J. C.; Lu, G. Y.; Quiocho, F. A. *J. Biol. Chem.* **1991**, *266*, 5202-5219.

(30) Rohrbach, M. R.; Braun, V.; Koster, W. *J. Bacteriol.* **1995**, *177*, 7186-7193.

(31) Boulanger, P.; leMaire, M.; Bonhivers, M.; Dubois, S.; Desmadril, M.; Letellier, L. *Biochemistry-US* **1996**, *35*, 14216-14224.

(32) Sprencel, C.; Cao, Z.; Qi, Z.; Scott, D. C.; Montaugue, M. A.; Ivanoff, N.; Xu, J.; Raymond, K. N.; Newton, S. M. C.; Klebba, P. E. *J. Bacteriol.* **2000**, *182*, 5359-5364.

(33) Rees, D. C.; Lee, A. T.; Borths, E. L.; Locher, K. P. *Faseb J.* **2003**, *17*, A1185-A1185.

(34) Miethke, M.; Klotz, O.; Linne, U.; May, J. J.; Beckering, C. L.; Marahiel, M. A. *Mol. Microbiol.* **2006**, *61*, 1413-1427.

(35) Benach, J.; Neely, H.; Seethataman, J.; Chen, C. X.; K., C.; Ma, L.-C.; Janjua, H.; Xiao, R.; Baran, M.; Acton, T. B. *et al.* Protein Data Bank accession number 2PHZ **2007**, *unpublished work*.

(36) Müller, A.; Wilkinson, A. J.; Wilson, K. S.; Duhme-Klair, A. K. *Angew. Chem. Intl. Ed.* **2006**, *45*, 1673-1685.

(37) Park, S. F.; Richardson, P. T. *J. Bacteriol.* **1995**, *177*, 2259-2264.

(38) Sebulsky, M. T.; Shilton, B. H.; Speziali, C. D.; Heinrichs, D. E. *J. Biol. Chem.* **2003**, *278*, 49890-49900.

(39) Butterton, J. R.; Calderwood, S. B. *J. Bacteriol.* **1994**, *176*, 5631-5638.

(40) Protein Data Bank accession number 2GPJ, *unpublished work*.

(41) Matzanke, B. F.; Anemuller, S.; Schunemann, V.; Trautwein, A. X.; Hantke, K. *Biochemistry-US* **2004**, *43*, 1386-1392.

(42) Brickman, T. J.; McIntosh, M. A. *J. Biol. Chem.* **1992**, *267*, 12350-12355.

(43) Roosenberg, J. M.; Lin, Y. M.; Lu, Y.; Miller, M. J. *Curr. Med. Chem.* **2000**, *7*, 159-197.

(44) Braun, V.; Pramanik, A.; Gwinner, T.; Köberle, M.; Bohn, E. *Biometals* **2009**, *22*, 3-13.

(45) Abergel, R. J.; Wilson, M. K.; Arceneaux, J. E. L.; Hoette, T. M.; Strong, R. K.; Byers, B. R.; Raymond, K. N. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18499-18503.

(46) Cendrowski, S.; MacArthur, W.; Hanna, P. *Mol. Microbiol.* **2004**, *51*, 407-417.

(47) Lee, J. Y.; Janes, B. K.; Passalacqua, K. D.; Pfleger, B. F.; Bergman, N. H.; Liu, H.; Håkansson, K.; Somu, R. V.; Aldrich, C. C.; Cendrowski, S. *et al. J. Bacteriol.* **2007**, *189*.

(48) Oves-Costales, D.; Kadi, N.; Fogg, M. J.; Song, L.; Wilson, K. S.; Challis, G. L. *J. Am. Chem. Soc.* **2007**, *129*, 8416-.

(49) Zawadza, A.; Abergel, R. J.; Nichiporuk, R.; Andersen, U. J.; Raymond, K. N. *Biochemistry-US* **2009**, *48*, 3645-3657.

(50) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1541-1546.

(51) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D. *et al. Nature* **2002**, *417*, 141-147.

(52) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nature Chem. Biol.* **2005**, *1*, 265-269.

(53) Barona-Gomez, F.; Lautru, S.; Francou, F. X.; Leblond, P.; Pernodet, J. L.; Challis, G. L. *Microbiol.- Sgm* **2006**, *152*, 3355-3366.

(54) Barona-Gomez, F.; Wong, U.; Giannakopulos, A. E.; Derrick, P. J.; Challis, G. L. *J. Am. Chem. Soc.* **2004**, *126*, 16282.

(55) Kadi, N.; Oves-Costales, D.; Barona-Gomez, F.; Challis, G. L. *Nature Chem. Biol.* **2007**, *3*, 652-656.

(56) Tunca, S.; Barreiro, C.; Sola-Landa, A.; Coque, J. J. R.; Martin, J. F. *FEBS J.* **2007**, *274*, 1110-1122.

(57) Challis, G. L.; Ravel, J. *FEMS Microbiol. Lett.* **2000**, *187*, 111-114.

(58) Lautru, S.; Oves-Costales, D.; Pernodet, J. L.; Challis, G. L. *Microbiol.- Sgm* **2007**, *153*, 1405–1412.

(59) Bunet, R.; Brock, A.; Rexer, H.-U.; Takano, E. *FEMS Microbiol. Lett.* **2006**.

(60) Kieser, T.; Chater, K. F.; Bibb, M. J.; Hopwood, D. A. *Practical Streptomyces Genomics 1st Edition*; The John Innes Foundation: Norwich, UK, 2000.

(61) Clancy, A.; Loar, J. W.; Speziali, C. D.; Oberg, M.; Heinrichs, D. E.; Rubens, C. E. *Mol. Microbiol.* **2006**, *59*, 707-721.

(62) Kim, D.-W.; Chater, K. F.; Lee, K.-J.; Hesketh, A. *Microbiology* **2005**, *151*, 2707-2720.

(63) Widdick, D. A.; Dilks, K.; Chandra, G.; Bottrill, A.; Naldrett, M.; Pohlschroder, M.; Palmer, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17927-17932.

(64) Robinson, C. *Biol. Chem.* **2000**, *381*, 89.

(65) Hutchings, M. I.; Palmer, T.; Harrington, D. J.; Sutcliffe, I. C. *Cell Press* **2008**, *17*, 13-21.

(66) Wayne, R.; Frick, K.; Neilands, J. B. *J. Bacteriol.* **1976**, *126*, 7-12.

(67) Huschka, H.; Naegeli, H. U.; Leuenbergerryf, H.; Kellerschierlein, W.; Winkelmann, G. *J. Bacteriol.* **1985**, *162*, 715-721.

(68) Hossain, M. B.; Jalal, M. A. F.; Benson, B. A.; Barnes, C. L.; Vanderhelm, D. *J. Am. Chem. Soc.* **1987**, *109*, 4948-4954.

(69) Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. *Science* **1998**, *282*, 2215-2220.

(70) Sauer, M.; Hantke, K.; Braun, V. *Mol. Microbiol.* **1990**, *4*, 427-437.

(71) Failes, T. W.; Hambley, T. W. *Aust. J. Chem.* **2000**, *53*, 879.

(72) Dhungana, S.; White, P. S.; Crumbliss, A. L. *J. Biol. Inorg. Chem.* **2001**, *6*, 810.

(73) Deeth, R. J.; Fey, N.; Williams-Hubbard, B. *J. Comput. Chem.* **2005**, *26*, 123-30.

(74) MOE, 2005, Chemical Computational Group, Montréal, Canada.

(75) Stephan, H.; Freund, S.; Meyer, J. M.; Winkelmann, G.; Jung, G. *Liebigs. Ann. Chem.* **1993**, 43-48.

(76) Kant, C., University of Warwick, 2005.

(77) Gräslund, S.; Nordlund, P.; Weigelt, J.; Bray, J.; Gileadi, O.; Knapp, S. *Nature Met.* **2008**, *5*, 135-146.

(78) Invitrogen 2004.

(79) Clone Manager Suite 7.0.1.1, Scientific and Educational Software

(80) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, 2001.

(81) Bradford, M. M. *Anal. Biochem.* **1976**, *72*, 248-254.

(82) Kapust, R. B.; Tözsér, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S. *Prot. Eng.* **2001**, *14*, 993-1000.

(83) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry - Part II: Techniques for the study of biological structure and function*; W. H. Freeman & Co., 1980.

(84) Miller, D. M. d.; Olson, J. S.; Pflugrath, J. W.; Quiocho, F. A. *J. Biol. Chem.* **1983**, *258*, 13665-13672.

(85) Ahnström, J.; Faber, K.; Axler, O.; Dahlbäck, B. *J. Lipid. Res.* **2007**, *48*, 1754.

(86) Origin 6.0 Professional, Microcal Software Inc.

(87) Higgins, C. F. *Nature* **2007**, *446*, 749-757.

(88) Vertesy, L.; Aretz, W.; Fehlhaber, H. W.; Kogler, H. *Helv. Chim. Acta.* **1995**, *78*, 46-60.

(89) Krewulak, K. D.; Shepherd, C. M.; Vogel, H. J. *Biometals* **2005**, *18*, 375-386.

(90) Rodger, A., *personal communication.*

(91) Shi, R.; Proteau, A.; Wagner, J.; Cui, Q.; Purisima, E. O.; Matte, A.; Cygler, M. *Proteins* **2008**, *75*, 598-605.

(92) Dhungana, S.; White, D. J.; Crumbliss, A. L. *J. Am. Chem. Soc.* **2003**, *125*, 14760.

(93) Dietrich, A.; Fidelis, K. A.; Powell, D. R.; van der Helm, D.; Eng-Wilmot, D. L. *J. Chem. Soc. Dalton. Trans.* **1991**, 231.

(94) Hou, Z.; Sunderland, C. J.; Nishio, T.; Raymond, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 5148.

(95) Kubow, S. A.; Takeuchi, K. J.; Gyzbowski, J. J.; Jircitano, A. J.; Goedken, V. L. *Inorg. Chim. Acta.* **1996**, *241*, 21-30.

(96) Kerpert, D. L. *Inorg. Chem.* **1972**, *11*, 1561-1563.

(97) Jensen, F. *Introduction to Computational Chemistry*; 2nd ed.; Wiley, 2007.

(98) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490-519.

(99) Allured, V. S.; Kelly, C.; Landis, C. R. *J. Am. Chem. Soc.* **1991**, *113*, 1-12.

(100) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1998**, *120*, 2641-2649.

(101) Comba, P.; Hambley, T. W. *Molecular Modelling of Coordination Compounds*; VCH, 1995.

(102) Deeth, R. J. *Faraday Discuss* **2003**, *124*, 379-91.

(103) Zinelabidine, A.; Bouraoui, A.; Mhenni, F.; Blaive, B.; Gallo, R. *J. Mol. Struct. - Theochem* **1993**, *105*, 267-274.

(104) van der Helm, D.; Polling, M. *J. Am. Chem. Soc.* **1976**, *98*, 82.

(105) Herrera, J. O. M.; Coelho Paes, L.; Canavarro Benites, A. M.; Machado, R. M. e. S. P. *Quim. Nova* **2006**, *29*, 269-276.

(106) Edwards, D. C.; Myneni, S. C. B. *J. Phys. Chem.* **2005**, *109*, 10249-10256.

(107) Felder, C. E.; Shanzer, A. *Biopolymers* **2003**, *68*, 407-421.

(108) Jaguar v3.5, 1998, Schrödinger, Inc., Portland, OR.

(109) Thönnes, E. □Topics in Statistics lecture notes,□ University of Warwick, 2003.

(110) Goffe, B.; Ferrier; Rogers *J. Econom.* **1994**, *60*, 65-100.

(111) Nelder, J. A.; Mead, R. *Computer Journal* **1965**, *7*, 308-313.

(112) Barnes, C. L.; Eng-Wilmot, D. L.; van der Helm, D. *Sect.C:Cryst.Struct.Commun* **1984**, *40*, 922.

(113) Hossain, M. B.; Jalal, M. A. F.; Benson, B. A.; Barnes, C. L.; van der Helm, D. *J. Am. Chem. Soc.* **1987**, *109*, 4948.

(114) Amsterdam Density Functional, 2005, Scientific Computing and Modelling NV, Amsterdam.

(115) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244-13249.

APPENDICES

A.1. The estimation of F_{min} by previous studies

A.1.1 Simulated quenching data

To test the different methods used to estimate the value of F_{min}, two "perfect" datasets were created such that their fluorescence maximum and minimum (F_0 and F_{min}) were set to 400 and 300 respectively, with different dissociation constants, (K_D = 0.2, 2.0 μM) using the one site binding model:

$$\frac{F_0 - F}{F_0 - F_{\min}} = \frac{K_D + [P_T] + [L_T] - \sqrt{((K_D + [P_T] + [L_T])^2 - 4[P_T][L_T])}}{2[P_T]} \quad \ldots(1)$$

where F is the fluorescence of the hypothetical solution in the presence of $[L_T]$ concentration of ligand. The protein concentration was fixed at $[P_T]$ = 2.257 μM and $[L_T]$ varied from 0 – 30 μM. The datasets (F versus $[L_T]$) are illustrated in Figure 0-1.

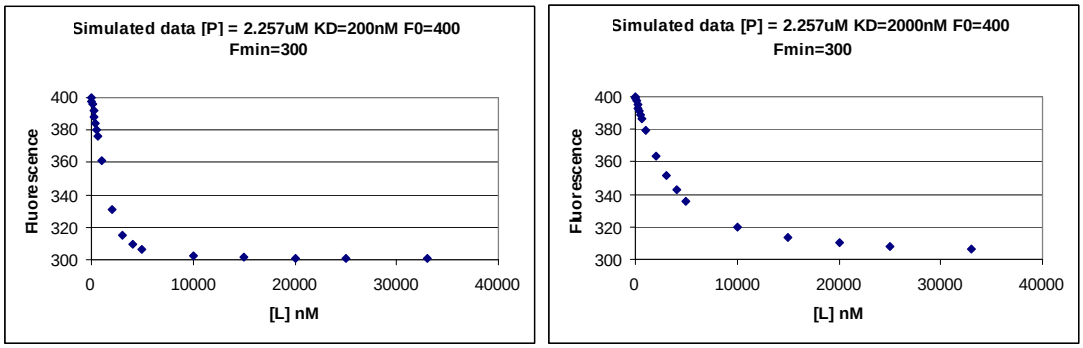


Figure 0-1: Graphs of simulated datasets of a ligand L binding to a fluorophore P where F_0=400, F_{min}=300, [P] = 2.257 μM, with (left) K_D = 200 nM, (right) K_D = 2.0 μM

A.1.2 Method of Rohrbach *et al.*

In Rohrbach *et al.*[30] the maximal quenching of the fluorescence of the *E. coli* protein FhuD was determined "from the intercept of the straight line obtained by double-reciprocal plotting of the substrate concentration versus the fluorescence quenching", with the x-axis being the reciprocal of the fluorescence quenching. There was no explanation or reference to the derivation of this method.

Using this method on the two datasets of known F_{min} (Figure 0-2), these gave estimations of the intercept as 3.0876 and 3.3348 for the simulated data with K_D = 0.2 μM and K_D = 2 μM respectively. To convert the intercept into estimates of F_{min}, the unknown F was found in the expression (3.0876 = $F_0/(F_0 - F)$) which corresponded to F_{min} values of 270 and 280 respectively, which were not close to the true value of 300.

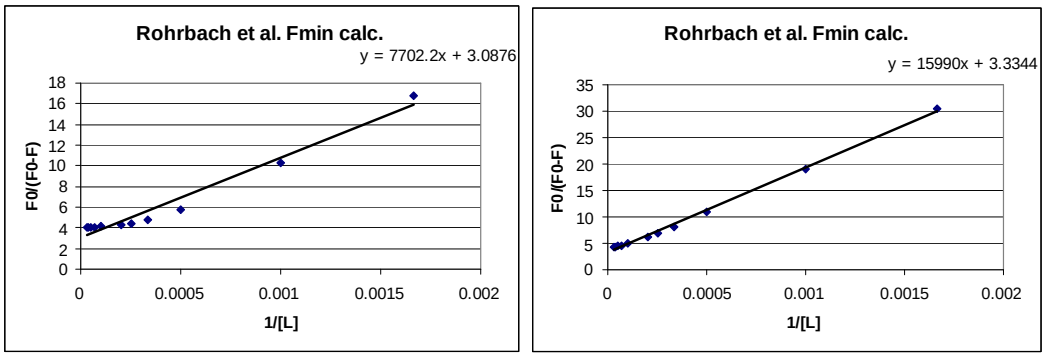


Figure 0-2: The estimation of F_{min} from the simulated data using the method of Rohrbach *et al.*[30]

The intercept of these lines are supposed to give the value of F_{min} after linear transformation (the value of 4 would correspond to F_{min} = 300) but from the calculated intercepts of the lines from the equations on the top-right corner they are far out. (left) K_D = 200 nM, (right) K_D = 2000nM.

A.1.3 Method of Sebulsky *et al.*

Sebulsky *et al.*[38] have suggested the equation to estimate the value of F_{min}:

$$\frac{[L_T]}{[L_T]+c} = \frac{1}{F_{\min}}\left(\frac{F_0 - F}{F_0}\right) \ldots(2)$$

$[L_T]$ is the total concentration of ligand and c is a constant. Again, there was no explanation or reference to the derivation of this relation. To estimate F_{min} from this relation, one assumes one has to plot $[L_T]/([L_T]+c)$ against $(F_0-F)/F_0$ with an arbitrary c, and the resulting gradient of the line will be $1/F_{min}$, but the transformation of the simulated datasets did not give a straight line and the gradient of this curve varied greatly (Figure 0-3).

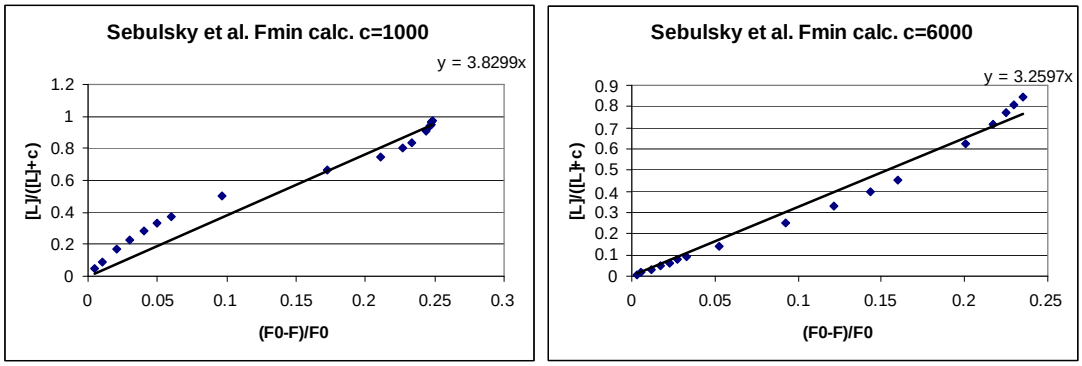


Figure 0-3: The estimation of F_{min} from the simulated data using the method of Sebulsky *et al.*

The left hand side of Equation 2 was plotted against the right hand side, and *c* was chosen such that the curve was roughly linear. The gradient 4 would give the desired value of F_{min} after linear transformation. (left) K_D = 0.2 µM (right) K_D = 2 µM

Double reciprocal plotting of these curves does give an almost linear curve (Figure 0-4), but this time the gradient gives a value of F_{min} (after transformation) and the value of c was adjusted to give a relationship which was nearly of the form $y = mx$ (a line with intercept zero), where the relationship is analogous to $\frac{[L_T]+c}{[L_T]} = F_{\min}\left(\frac{F_0}{F_0 - F}\right)$. The gradients of these graphs correspond to F_{min} values of 193 and 258 for the simulated data with K_D = 0.2 µM and K_D = 2 µM respectively, which were inaccurate estimates of the true value of 300.

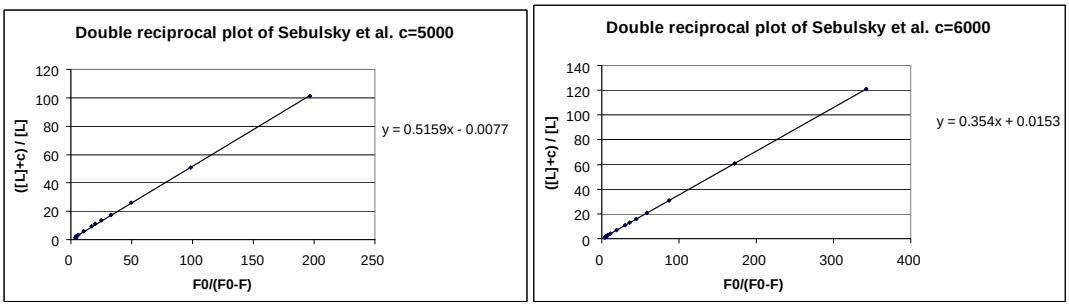


Figure 0-4: The estimation of F_{min} from the simulated data using the method of Sebulsky *et al.* - double-reciprocal plots.

The reciprocal of the left hand side of Equation 2 was plotted against the reciprocal of the right hand side, and *c* was chosen such that the curve had an intercept of nearly zero, such that the gradient corresponded directly to F_{min} after linear transformation. A gradient of 0.25 would correspond to F_{min} = 300, which both do not (left) K_D = 0.2 µM, (right) K_D = 2 µM.

A.2. Derivation of one site binding equation

Assume the binding of protein to the substrate is one-to-one, i.e.:

$$P + S \leftrightarrow PS \quad \ldots(1)$$

The dissociation constant K_D is given by the formula:

$$K_D = \frac{[P][S]}{[PS]} \ldots(2)$$

But both the free concentrations of protein and substrates, [P] and [S] respectively, can be expressed as the sum of the total concentration $[P_T]$ and $[S_T]$ minus the concentration of bound species [PS]:

$$\begin{aligned} [P] &= [P_T] - [PS] \\ [S] &= [S_T] - [PS] \end{aligned} \quad \ldots(3)$$

Substituting into (2) and rearranging:

$$K_D = \frac{[P][S]}{[PS]} = \frac{([P_T] - [PS])([S_T] - [PS])}{[PS]}$$

$$[PS]\,K_D = [P_T][S_T] - ([P_T] + [S_T])[PS] + [PS]^2$$

$$[PS]^2 - (K_D + [P_T] + [S_T])[PS] + [P_T][S_T] = 0 \quad \ldots(4)$$

Equation (3) is now a quadratic equation in [PS], which can be easily solved by the quadratic formula:

$$[PS] = \frac{K_D + [P_T] + [S_T] \pm \sqrt{((K_D + [P_T] + [S_T])^2 - 4[P_T][S_T])}}{2} \quad \ldots(5)$$

But $[PS]/[P_T]$ is precisely the proportion of protein bound to the substrate, which can be monitored by techniques such as fluorescence quenching if the fluorescence of the non-bound F_0 and fully bound protein F_{min} is known:

Hence:

$$\frac{[PS]}{[P_T]} = \frac{F_0 - F}{F_0 - F_{min}} = \frac{K_D + [P_T] + [S_T] - \sqrt{((K_D + [P_T] + [S_T])^2 - 4[P_T][S_T])}}{2[P_T]} \ldots(6)$$

where F is the fluorescence of the protein at a given total protein $[P_T]$ and total substrate $[S_T]$ concentration.

A.3. Origin function files

A.3.1 Onesitebind2 method

```
[GENERAL INFORMATION]
Function Name=Onesitebind2
Brief Description=user4
Function Source=N/A
Function Type=User-Defined
Function Form=Y-Script
Number Of Parameters=2
Number Of Independent Variables=1
Number Of Dependent Variables=1

[FITTING PARAMETERS]
Naming Method=User-Defined
Names=Kd, P
Meanings=Dissociation Constant, Protein concentration
Initial Values=500(V), 2000(V)
Lower Bounds=0(X, On),0(X, On)
Upper Bounds=--(X, Off),--(X, Off)
Number Of Significant Digits=

[FORMULA]
q = Kd+x+P;
y = ( q - ( q^2 - (4*x*P) ) ^ 0.5 ) / (2*P)
[CONSTRAINTS]
/*Enter general linear constraints here*/
[CONSTANTS]

[INITIALIZATIONS]
/*Scripts to be executed before fitting, a good place for
complicated initialization.*/
[AFTER FITTING]
/*Scripts to be executed after fitting, a good place for generating
results.*/

[INDEPENDENT VARIABLES]
x=
[DEPENDENT VARIABLES]
y=

[CONTROLS]
General Linear Constraints=Off
Initialization Scripts=Off
Scripts After Fitting=Off
Number Of Duplicates=N/A
Duplicate Offset=N/A
Duplicate Unit=N/A
Generate Curves After Fitting=Yes
Curve Point Spacing=Same X as Fitting Data
Generate Peaks After Fitting=Yes
Generate Peaks During Fitting=Yes
Generate Peaks with Baseline=Yes
Paste Parameters to Plot After Fitting=Yes
Paste Parameters to Notes Window After Fitting=Yes
Generate Residue After Fitting=No
Keep Parameters=No
```

A.3.2 Onesitebind_lipid

```
[GENERAL INFORMATION]
Function Name=Onesidebind_lipid
Brief Description=user5
Function Source=N/A
Function Type=User-Defined
Function Form=Y-Script
Number Of Parameters=3
Number Of Independent Variables=1
Number Of Dependent Variables=1

[FITTING PARAMETERS]
Naming Method=User-Defined
Names=Kd, P, IPL
Meanings=?
Initial Values=500(V),2000(V),1(V)
Lower Bounds=--(X,OFF)
Upper Bounds=--(X,OFF)
Number Of Significant Digits=

[FORMULA]
q = (x - Kd - P)/2;
Lf = q + (q^2 + Kd * x )^0.5;
qq = Lf / (Lf + Kd);
y = IPL * qq

[CONSTRAINTS]
/*Enter general linear constraints here*/
[CONSTANTS]

[INITIALIZATIONS]
/*Scripts to be executed before fitting, a good place for
complicated initialization.*/
[AFTER FITTING]
/*Scripts to be executed after fitting, a good place for generating
results.*/

[INDEPENDENT VARIABLES]
x=
[DEPENDENT VARIABLES]
y=

[CONTROLS]
General Linear Constraints=Off
Initialization Scripts=Off
Scripts After Fitting=Off
Number Of Duplicates=N/A
Duplicate Offset=N/A
Duplicate Unit=N/A
Generate Curves After Fitting=Yes
Curve Point Spacing=Same X as Fitting Data
Generate Peaks After Fitting=Yes
Generate Peaks During Fitting=Yes
Generate Peaks with Baseline=Yes
Paste Parameters to Plot After Fitting=Yes
Paste Parameters to Notes Window After Fitting=Yes
Generate Residue After Fitting=No
Keep Parameters=No
```

A.4. Dissociation constant calculation graphs

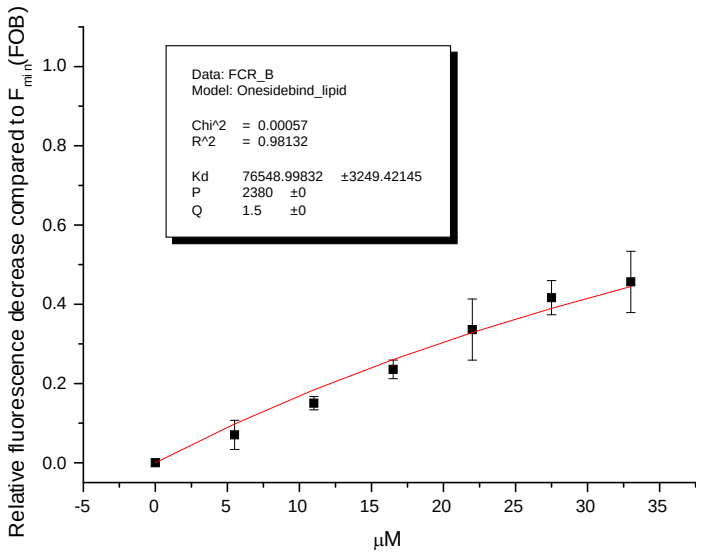


Figure 0-5: The relative fluorescence of ferrichrome decrease compared to F_{min} of ferrioxamine B bound to CdtB to a 78 µg/mL CdtB solution at 21 °C.

The model was fitted to the Onesitebind_lipid method, and the non-linear curve fitting function was used to find a suitable K_D and Q given [P] = 2.380 µM. The best fit to the data was obtained with K_D = 77 µM.

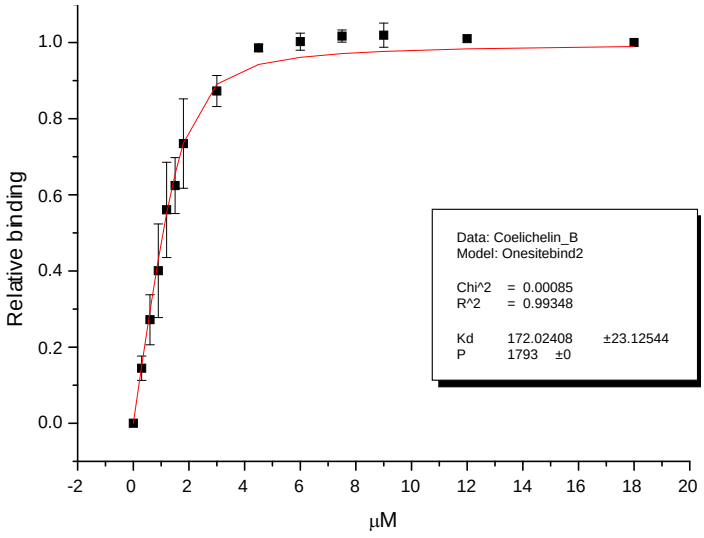


Figure 0-6: The relative binding of ferri-coelichelin to a 78 µg/mL CdtB solution at 21 °C on assumption that 75% of the protein participates in binding.

The model was fitted to the Onesitebind2 method, and the non-linear curve fitting function was used to find a suitable K_D given [P] = 1.790 µM. The best fit to the data was obtained with K_D = 172 nM.

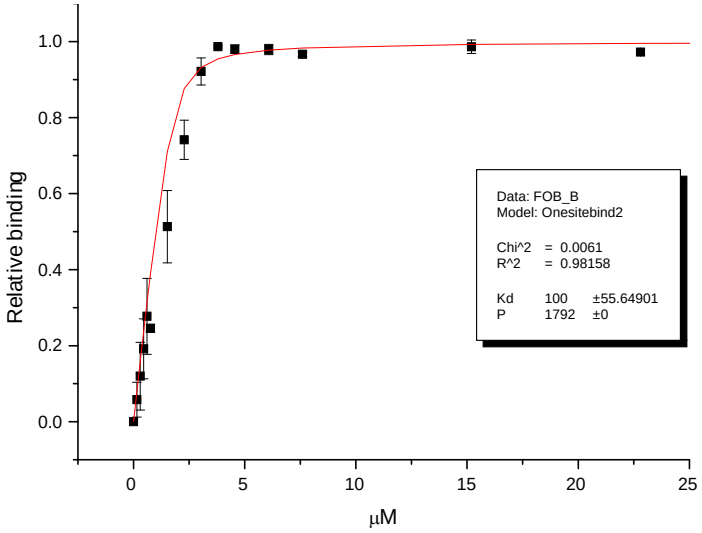


Figure 0-7: The relative binding of ferrioxamine B to a 78 mg/mL CdtB solution at 21 °C on assumption that 75% of the protein participates in binding.

The model was fitted to the Onesitebind2 method, and the non-linear curve fitting function was used to find a suitable K_D given [P] = 1.790 μM. The fit to the data was obtained with K_D = 100 nM.

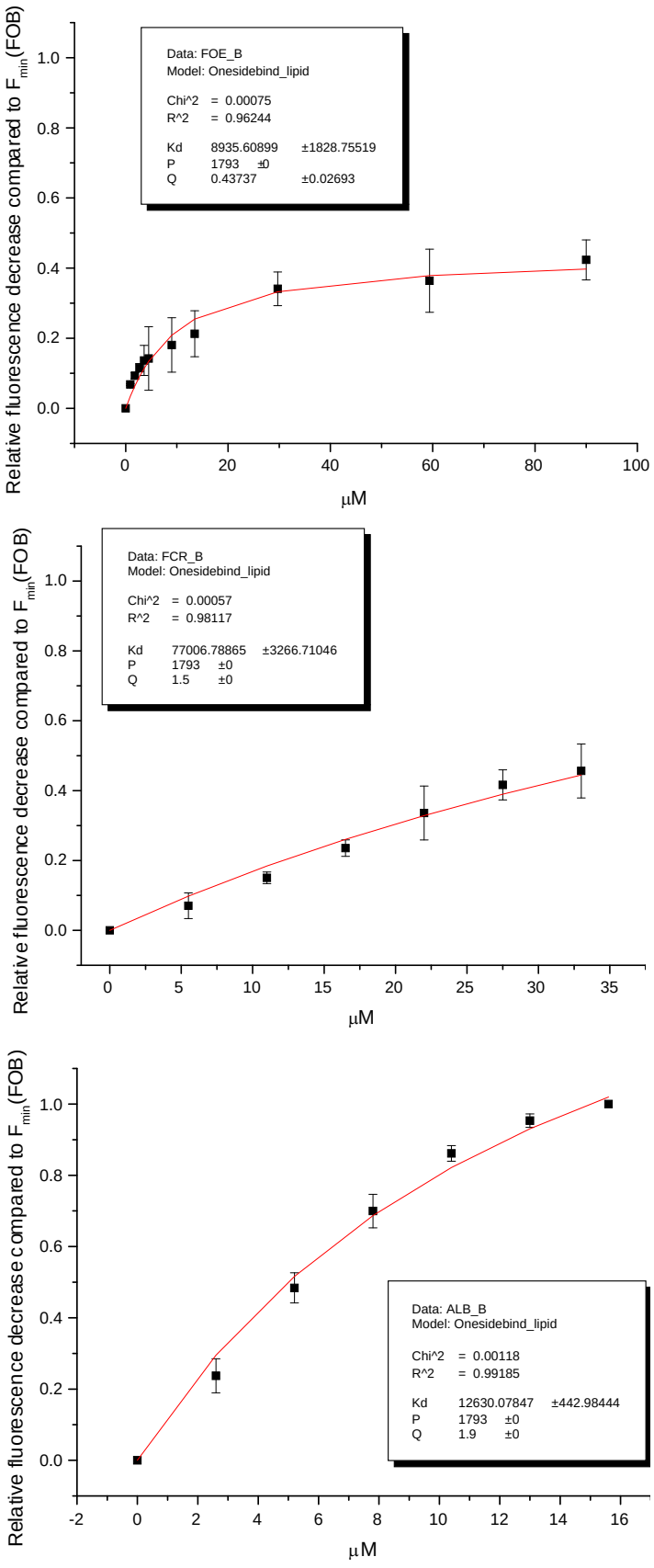


Figure 0-8: The relative fluorescence decrease on addition of ferri-siderophores of a 78 µg/mL CdtB solution at 21 °C compared to F_{min} of ferrioxamine B bound to CdtB, on assumption 75% of the protein can participate in binding.

The model was fitted to the Onesitebind_lipid method, and the non-linear curve fitting function was used to find a suitable K_D and Q given [P] = 1.793 µM. (Top) ferrioxamine E; (middle) ferrichrome; (bottom) ferri-albomycin.

A.5. Modelling of ferric hydroxamates

Table 0-1: Summary of the 27 X-ray crystal structures of ferric-*tris*-hydroxamates used in the survey of bond lengths and angles of ferric-*tris*-hydroxamates.

The CDS reference code is the code used by the Cambridge Crystallographic Database at Daresbury to refer to the crystal structure, with the common name or chemical name and the original reference also given. Only one ferric iron centre was included in the multinuclear complexes IRADEK and TEQKUV.

CDS Reference code	Common name	Reference
BOPMUO01	Asperchrome A	M.B.Hossain, D.van der Helm, M.Alam, R.Sanduja (1983) ACA,Ser.2 ,**11**,29
BUHQUQ	(tris((3-(N-Acetyl-N-hydroxy)glycylamino) propyl)ammonium)-iron(iii)	K.Matsumoto, N.Suzuki, T.Ozawa, K.Jitsukawa, H.Masuda (2001) *Eur.J.Inorg.Chem.* ,2481
CETWIH	Ferric neurosporin	D.L.Eng-Wilmot, A.Rahman, J.V.Mendenhall, S.L.Grayson, D.van der Helm (1984) *J.Am.Chem.Soc.* ,**106,**1285
COFDIK10	Neocoprogen I	M.B.Hossain, M.A.F.Jalal, B.A.Benson, C.L.Barnes, D.van der Helm (1987) *J.Am.Chem.Soc.* ,109,4948
CUJHIY	Ferrirubin hydrate	C.L Barnes, M. B. Hossain, M.A.F.Jalal, B.A.Benson, D.L. Eng-Wilmot, S.L Grayson, S.K.Agarwal, R.Mocheria, D.van der Helm (1985) *Acta Crystallogr.,Sect.C:Cryst.Struct.Commun.* **41**, 341
DEWNAU	tris(μ_2-1.5-bis((1,2-Dihydro-1-hydroxy-2-oxopyridin-6-yl)carbonyl) -5-diazapentane-di-iron	R.C.Scarrow, D.L.White, K.N.Raymond (1985) *J.Am.Chem.Soc.* **107**, 6540
DUPJON	Ferrioxamine D1	M.B.Hossain, M.A.F.Jalal, D.van der Helm (1986) *Acta Crystallogr.,Sect.C:Cryst.Struct.Commun.* ,42,1305
FEBOAH02	tris(Benzohydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,53,879
FERCRN10	Ferricrocin	C.L.Barnes, D.L.Eng-Wilmot, D.van der Helm (1984) *Acta Crystallogr., Sect.C:Cryst.Struct.Commun.* ,**40,**922
FERMAH11	Ferrichrome-A	D.van der Helm, J.R.Baker, R.A.Loghry, J.D.Ekstrand (1981) *Acta Crystallogr.,Sect.B:Struct.Crystallogr.Cryst.Chem.* ,**37,**323
FEROXE10	Ferrioxamine E	D.van der Helm, M.Poling (1976) J.*Am.Chem.Soc.* ,**98,**82
FERRID10	Ferrichrome	D.van der Helm, J.R.Baker, D.L.Eng-Wilmot, M.B.Hossain, R.A.Loghry (1980) *J.Am.Chem.Soc.* ,**102,**4224
FTAFUS10	N,N',N''-Triacetylfusarinine-iron(iii)	M.B.Hossain, D.L.Eng-Wilmot, R.A.Loghry, D.van der Helm (1980) *J.Am.Chem.Soc.* ,**102,**5766
IRADEK	(μ_2-1,3-Benzodihydroxamato-O-O')-tetra-iron	Y.Bai, D.Guo, C-y Duan, D-b Dang, K-L Pang, Q-J Meng (2004) *Chem. Commun.* 186
JEPRAX	Ferrirhodin	K.Fidelis, M.B.Hossain, M.A.F.Jajal, D.van der Helm (1990) *Acta Crystallogr., Sect.C:Cryst.Struct.Commun.* ,**46,**1612
OCELOX	(tris[2-(((N-acetyl-N-oxy)glycylamino)ethyl)amine])-iron(iii)	K.Matsumoto, T.Ozawa, K.Jitsukawa, H.Einaga, H.Masuda (2001) *Inorg.Chem.* ,**40,**190
OFUYET	Ferrioxamine B	S.Dhungana, P.S.White, A.L.Crumbliss (2001) *J.Biol.Inorg.Chem.(JBIC)* ,**6,**810
QOZDUE	Λ-(((R,R,R)-tris(2-((N-Acetyl-N-hydroxy) -D-alanylamino) ethyl) amine)- iron(iii))	K.Matsumoto, T.Ozawa, K.Jitsukawa, H.Einaga, H.Masuda (2001) *Chem.Commun.* ,978
SIPMEJ	Retro-ferrioxamine E	M.B.Hossain, M.A.F.Jalal, D.van der Helm, K.Shimizu, M.Akiyama (1998) *J.Chem.Cryst.* ,**28,**53
SISSAO	tris(N-(p-Tolyl) acetohydroxamato-	A.Dietrich, K.A.Fidelis, D.R.Powell, D.van der

	O,O')-iron(iii) acetone solvate	Helm, D.L.Eng-Wilmot (1991) J.*Chem.Soc.,Dalton Trans.*,231
SISSES	tris(N-Methyl-4-methylbenzohydroxamato-O,O')-iron(iii)	A.Dietrich, K.A.Fidelis, D.R.Powell, D.van der Helm, D.L.Eng-Wilmot (1991) J.*Chem.Soc.,Dalton Trans.* ,231
SUXREI (fac)	*fac*-tris(Hydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,**53,**879
SUXREI (mer)	*mer*-tris(Hydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,**53,**879
TEQKUV	(μ_2-Alcaligin)-bis(alcaligin)-di-iron	Hou, Z.; Sunderland, C. J.; Nishio, T.; Raymond, K. N., *J.Am. Chem. Soc.* 1996, **118**, 5148.
UNEVUE	bis(Ferrioxamine B) bis (benzo-18-crown-6)	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760
UNEWAL	bis(Ferrioxamine B) bis (cis-syn-cis-dicyclohexano-18-crown-6)	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760
UNEWEP	bis(Ferrioxamine B) 18-crown-6 ether	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760

Table 0-2: Summary of the 23 X-ray crystal structures of ferric-*tris*-hydroxamates used in the DOMMIMOE parameter refinement.

CDS Reference code	Common name	Reference
BOPMUO01	Asperchrome A	M.B.Hossain, D.van der Helm, M.Alam, R.Sanduja (1983) ACA,Ser.2 ,11,29
BUHQUQ	(tris((3-(N-Acetyl-N-hydroxy)glycylamino) propyl)ammonium)-iron(iii)	K.Matsumoto, N.Suzuki, T.Ozawa, K.Jitsukawa, H.Masuda (2001) *Eur.J.Inorg.Chem.* ,2481
CETWIH	Ferric neurosporin	D.L.Eng-Wilmot, A.Rahman, J.V.Mendenhall, S.L.Grayson, D.van der Helm (1984) *J.Am.Chem.Soc.* ,**106,**1285
COFDIK10	Neocoprogen I	M.B.Hossain, M.A.F.Jalal, B.A.Benson, C.L.Barnes, D.van der Helm (1987) *J.Am.Chem.Soc.* ,**109,**4948
DUPJON	Ferrioxamine D1	M.B.Hossain, M.A.F.Jalal, D.van der Helm (1986) *Acta Crystallogr.,Sect.C:Cryst.Struct.Commun.* ,**42,**1305
FEBOAH02	tris(Benzohydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,**53,**879
FERCRN10	Ferricrocin	C.L.Barnes, D.L.Eng-Wilmot, D.van der Helm (1984) *Acta Crystallogr., Sect.C:Cryst.Struct.Commun.* ,**40,**922
FERMAH11	Ferrichrome-A	D.van der Helm, J.R.Baker, R.A.Loghry, J.D.Ekstrand (1981) *Acta Crystallogr.,Sect.B:Struct.Crystallogr.Cryst.Chem.* ,**37,**323
FEROXE10	Ferrioxamine E	D.van der Helm, M.Poling (1976) J.*Am.Chem.Soc.* ,**98,**82
FERRID10	Ferrichrome	D.van der Helm, J.R.Baker, D.L.Eng-Wilmot, M.B.Hossain, R.A.Loghry (1980) *J.Am.Chem.Soc.* ,**102,**4224
FTAFUS10	N,N',N''-Triacetylfusarinine-iron(iii)	M.B.Hossain, D.L.Eng-Wilmot, R.A.Loghry, D.van der Helm (1980) *J.Am.Chem.Soc.* ,**102,**5766
JEPRAX	Ferrirhodin	K.Fidelis, M.B.Hossain, M.A.F.Jajal, D.van der Helm (1990) *Acta Crystallogr., Sect.C:Cryst.Struct.Commun.* ,**46,**1612
OCELOX	(tris[2-(((N-acetyl-N-oxy)glycylamino)ethyl)amine])-iron(iii)	K.Matsumoto, T.Ozawa, K.Jitsukawa, H.Einaga, H.Masuda (2001) *Inorg.Chem.* ,**40,**190
OFUYET	Ferrioxamine B	S.Dhungana, P.S.White, A.L.Crumbliss (2001) *J.Biol.Inorg.Chem.(JBIC)* ,**6,**810
QOZDUE	Λ-(((R,R,R)-tris(2-((N-Acetyl-N-hydroxy) -D-alanylamino) ethyl) amine)- iron(iii))	K.Matsumoto, T.Ozawa, K.Jitsukawa, H.Einaga, H.Masuda (2001) *Chem.Commun.* ,978
SIPMEJ	Retro-ferrioxamine E	M.B.Hossain, M.A.F.Jalal, D.van der Helm, K.Shimizu, M.Akiyama (1998) *J.Chem.Cryst.* ,**28,**53

SISSAO	tris(N-(p-Tolyl) acetohydroxamato-O,O')-iron(iii) acetone solvate	A.Dietrich, K.A.Fidelis, D.R.Powell, D.van der Helm, D.L.Eng-Wilmot (1991) *J.Chem.Soc.,Dalton Trans.* ,231
SISSES	tris(N-Methyl-4-methylbenzohydroxamato-O,O')-iron(iii)	A.Dietrich, K.A.Fidelis, D.R.Powell, D.van der Helm, D.L.Eng-Wilmot (1991) *J.Chem.Soc.,Dalton Trans.* ,231
SUXREI (fac)	*fac*-tris(Hydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,**53,**879
SUXREI (mer)	*mer*-tris(Hydroxamato-O,O')-iron(iii)	T.W.Failes, T.W.Hambley (2000) *Aust.J.Chem.* ,**53,**879
UNEVUE	bis(Ferrioxamine B) bis (benzo-18-crown-6)	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760
UNEWAL	bis(Ferrioxamine B) bis (cis-syn-cis-dicyclohexano-18-crown-6)	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760
UNEWEP	bis(Ferrioxamine B) 18-crown-6 ether	S.Dhungana, P.S.White, A.L.Crumbliss (2003) *J.Am.Chem.Soc.* ,**125,**14760

Table 0-3: Performance of DOMMIMOE on X-ray crystallographic structures after parameter refinement.

The first column denotes the six-letter reference code of the structure in the Cambridge Crystallographic Database. Fe-O(C) and Fe-O(N) denote the Fe-oxygen bond lengths between carbonyl and oxime oxygens in the hydroxamate groups respectively. The RMSD The bite angle denotes the angle between the oxime oxygen–iron–carbonyl oxygen in a hydroxamate group in degrees. The calculated trigonal twist difference from the DOMMIMOE to the X-ray structures are given in the fifth column, and were calculated by the "eclipsing triangles" method.

CDS reference code	Fe-O(C) RMSD (Å)	Fe-O(N) RMSD (Å)	bite angle RMSD (°)	Trigonal twist difference (DOMMIMOE – X-ray) (°)
BOPMUO01	0.017	0.020	0.3	3.6
BUHQUQ	0.023	0.015	0.7	7.3
CETWIH	0.010	0.017	1.0	0.1
COFDIK10	0.022	0.029	1.5	0.8
DUPJON	0.007	0.018	0.4	5.3
FEBOAH02	0.015	0.005	0.7	12.2
FERCRN10	0.004	0.011	0.7	7.9
FERMAH11	0.025	0.014	1.0	8.3
FEROXE10	0.030	0.047	0.4	2.9
FERRID10	0.013	0.016	0.5	5.7
FTAFUS10	0.026	0.016	1.2	2.4
JEPRAX	0.017	0.020	1.0	1.7
OCELOX	0.016	0.020	0.6	6.6
OFUYET	0.014	0.011	0.3	8.6
QOZDUE	0.006	0.014	0.2	3.7
SIPMEJ	0.006	0.011	0.7	-4.2
SISSAO	0.024	0.007	1.3	12.5
SISSES	0.004	0.011	0.6	6.1
SUXREI (fac)	0.019	0.023	0.9	3.7

SUXREI (mer)	0.012	0.008	0.9	9.3
UNEVUE	0.021	0.010	0.7	7.2
UNEWAL	0.010	0.015	0.6	5.6
UNEWEP	0.020	0.009	0.7	7.2